

Reaping the Benefits of Bundling under High Production Costs

Will Ma*

David Simchi-Levi†

September 14, 2016

Abstract

It is well-known that selling different goods in a single bundle can significantly increase revenue, even when the valuations for the goods are independent. However, bundling is no longer profitable if the goods have high production costs. To overcome this challenge, we introduce a new mechanism, Pure Bundling with Disposal for Cost (PBDC), where after buying the bundle, the customer is allowed to return any subset of goods for their production cost. We derive both distribution-dependent and distribution-free guarantees on its profitability, which improve previous techniques. Our distribution-dependent bound suggests that the firm should never price the bundle such that the profit margin is less than $1/3$ of the expected welfare, while also showing that PBDC is optimal for a large number of independent goods. Our distribution-free bound suggests that on the distributions where PBDC performs worst, individual sales perform well. Finally, we conduct extensive simulations which confirm that PBDC outperforms other simple pricing schemes overall.

*Operations Research Center, Massachusetts Institute of Technology, willma@mit.edu.

†Institute for Data, Systems, and Society, Department of Civil and Environmental Engineering, and Operations Research Center, Massachusetts Institute of Technology, dslevi@mit.edu.

1 Introduction

We study the monopolist pricing problem of a firm selling n heterogeneous items. For each item, customers have a valuation, which is their maximum willingness-to-pay, drawn from a known distribution. A customer wants at most one of each item. The firm offers take-it-or-leave-it prices for every subset of items, and the customer chooses the subset maximizing her surplus (valuation for the subset minus price), with the no-purchase option always being available. We assume the customer’s valuation for a subset is additive over the items in the set. The objective of the firm is to maximize expected per-customer revenue.

In the full generality of the problem, the firm has $2^n - 1$ prices to set. However, it is important to find profitable yet *simple* pricing schemes that are determined by a small number of prices. Two such schemes are *Pure Components* (PC), where items are priced separately (and the price of a subset is understood to be the sum of its constituent prices), and *Pure Bundling* (PB), the strategy of only offering all the items together. A third scheme that generalizes both PC and PB is *Mixed Bundling* (MB), which offers individual item prices as well as a bundle price for all the items. MB can be seen as a form of price discrimination, where customers who highly value an item can buy it for its individual price, while customers with lower valuations still have a chance of buying it as part of a discounted bundle.

The efficacy of simple pricing schemes is of immense importance in retail, and has been studied over the past few decades in the economics literature, the operations research/marketing interface literature, and more recently, the computer science literature. For a single item, the solution is immediate: choose the price p maximizing $p(1 - F(p))$, where F is the CDF of the valuation (see [Mye81]). However, for two items, even if their valuations are independent, bundling can be better than individual sales.

For example, suppose we have two products with IID valuations, each of which is 1 half the time, and 2 half the time. If we sell the items individually, we can always get a sale for 1, or get a sale half the time for 2. In either case, the combined expected revenue is 2. However, if we sell the items as a bundle for 3, then the bundle will be purchased $\frac{3}{4}$ of the time, yielding an expected revenue of $\frac{9}{4}$.

The key observation is that the valuation of the bundle is more concentrated around its mean than the valuation of the individual items, which causes less consumer heterogeneity, and we can choose a price that is the highest willingness-to-pay for a larger fraction of customers. This makes it easier to reduce deadweight loss, which is revenue lost from pricing a customer with positive valuation out of the market, and consumer surplus, which is revenue lost from offering a customer a better price than necessary.

The power of bundling is even greater when valuations are negatively correlated—consider two products with marginal valuations that are uniform on $[0, 1]$ but correlated in a way such that they always sum to 1. In this case, offering the bundle at the price of 1 will always get a sale, extracting the entire welfare, while selling the items individually yields at most $\frac{1}{2}$, half the available surplus. These effects have long been known in the economics literature, following the pioneering work of

[Sti63], [AY76], [Sch84], and [MMW89].

Of course, bundling is not always superior to individual sales—this is especially true once we consider production costs. For example, suppose we have two goods with IID valuations that are uniform on $[0, 3]$, but each cost 2 to produce. Selling them individually at price $\frac{5}{2}$ will yield a profit of $\frac{1}{12}$ per item and is better than selling them as a bundle—these are low-profit-margin items that are only valuable to a small fraction of the population, and by bundling them we may force a customer into consuming a good for which her valuation is less than the production cost.

Over the decades, a lot of work has been done to compare the profit of Pure Bundling versus Pure Components. [AY76] write, “The chief defect of Pure Bundling is its difficulty in complying with Exclusion,” where Exclusion refers to the principle that a transfer is better off not occurring when the consumer’s valuation is below the producer’s cost. It is observed in [Sch84] for the case of bivariate normal valuations that Pure Bundling is better when mean valuations are high compared to costs. [BB99] prove that bundling a large number of goods can extract almost all of the available surplus, but this is crucially dependent on the items being “information goods”, i.e. goods with no production costs. [FN06] characterize conditions under which Pure Bundling outperforms Pure Components for a fixed number of items, and all of their conditions imply low costs. [LFCK13] define a measure of consumer heterogeneity that increases with costs, and present computational results showing Pure Bundling performs poorly relative to Pure Components as their measure of consumer heterogeneity increases.

The indisputable conclusion from all this work is that high costs are the greatest impediment to the magic of bundling. However, we argue that there is a simple way to enjoy the effects of bundling while allowing for the flexibility of individual sales—sell all of the items as a bundle, but allow the customer to return any subset of items for a refund equal to their total production cost. We call this scheme *Pure Bundling with Disposal for Cost* (PBDC). It is a strict improvement of Pure Bundling where the customer’s valuations that were below the cost have been *truncated* to equal the cost. Meanwhile, the firm is indifferent between producing an item for its cost or returning its cost to the customer, but PBDC makes it easier to sell the bundle because customers with low valuations for specific items won’t be priced out of the market.

Importantly, there is great flexibility in how to present PBDC to the customer in a transparent and attractive way. In fact, we show that PBDC has a few equivalent formulations which can already be seen in the market. One formulation is a *tariff* to enter the market, after which all products are sold at cost. Alternatively, PBDC can be introduced with an individual price for each item and a *per-item discount* for each item purchased beyond the first. From a marketing point of view, the tariff strategy is more attractive when the number of items is large, while the discount strategy is more attractive when the number of items is small.

Our scheme can be compared to that of [HC05], who recognized the need for a middle ground between Pure Bundling and Pure Components. They introduced the scheme *Customized Bundling*, which prices each bundle based only on its size, and not which items are included. [CLS08]¹ perform

¹See [CLS11] for the journal version.

extensive numerical experiments for the same scheme, calling it *Bundle-Size Pricing* (BSP), showing that it can extract 99% of the optimal profit in their simulations.

PBDC can be seen as orthogonal to BSP—while BSP imposes symmetric pricing across items but allows non-linear pricing based on quantity, PBDC allows asymmetric pricing across items based on cost but imposes additive pricing once the customer pays the tariff to enter the market. When all item costs are identical, PBDC is a simplified version of BSP, because instead of having n prices to decide, there is only one price to decide, be it thought of as the bundle price, the tariff, or the discount. However, since we are able to relate PBDC to Pure Bundling, it is much easier to analyze. Our work provides the first theoretical explanation for some of the empirical successes in [CLS08]—indeed, in their simulations, costs are either equal, or insignificant (equal to half of the product’s mean valuation).

We present two types of theoretical bounds in this paper. Both require that items have independent valuations, which is a standard and often necessary assumption in the bundling literature. Both also rely on a transformation from costs to negative valuations which as far as we know is new.

First, we prove that PBDC, with an appropriately chosen bundle price, extracts the entire welfare as the number of items approaches infinity, so long as the valuations have uniformly bounded variances. This type of result is based on the law of large numbers, which says that the sum of cost-truncated valuations, which we denote with the random variable X , lies within $(1 \pm \varepsilon)\mathbb{E}[X]$ with high probability. Therefore, the bundle price of $(1 - \varepsilon)\mathbb{E}[X]$ will be accepted by a $(1 - \varepsilon)$ -fraction of the customers, profiting almost the expected welfare, $\mathbb{E}[X] - C$ (C is the total cost of producing all the items), which is an upper bound on profit.

[BB99] have already proven this result for the case without costs, and [Arm99] has proven this result for a cost-based two-part tariff which we show is equivalent to PBDC, so this result in itself is not new. However, our analysis introduces the use of Cantelli’s stronger, one-sided concentration inequality in bundling, recovering previous bounds asymptotically and achieving a better bound when the coefficient of variation of X is large. In the latter case, both of the previous works recommend a bundle price of $C + \varepsilon(\mathbb{E}[X] - C)$, earning negligible profit, whereas our analysis never recommends a bundle price below $C + \frac{1}{3}(\mathbb{E}[X] - C)$ and guarantees a non-zero profit.

Furthermore, we recommend PBDC even when the number of items is small—the second type of theoretical bound we present is problem-independent, unaffected by the number of items or their variances. We prove that the expected profit of the best PBDC pricing is at least $\frac{1}{5.2}$ of the expected profit of the optimal mechanism, except in detectable pathological cases, where the best PC pricing achieves this guarantee. The benchmark in this theorem is the maximum expected profit that could be achieved from explicitly pricing all subsets of items². This is a tighter benchmark than expected welfare, which could be infinite without distributional assumptions.

We use tools from the computer science literature to bound this benchmark, most notably from [BILW14], who prove in the costless case that the better of PB and PC earns at least $\frac{1}{6}$ of

²Technically, the optimal mechanism is also allowed to offer lotteries of items, which have been shown to be necessary for achieving the optimum, in [HR12].

the optimal revenue. We improve their bound from $\frac{1}{6}$ to $\frac{1}{5.2}$ by using Cantelli’s inequality, and enhancing their *core-tail decomposition* technique in analyzing the core and tail together. We also construct an example improving the upper bound from $\frac{12}{13} \approx \frac{1}{1.08}$ to $\frac{3+\ln 2}{3+2\ln 2} \approx \frac{1}{1.19}$, where the previous best-known example is from [HN12]. Finally, we generalize the result of Babaioff et al. to the case with costs, where PBDC is needed instead of PB. We should point out that when the benchmark is the optimal mechanism, one cannot simply truncate all valuations from below by cost, because the optimal mechanism could exploit low valuations to reduce the cannibalization of higher-profit options. In general, profit is *non-monotone*, i.e. increasing customer valuations can decrease the optimal profit, as reported in [HN12].

In addition to the theoretical considerations, we provide a continuation of the numerical experiments from [CLS08], extensively testing the efficacy of PBDC on a finite number of items. We use the same independent demand distributions with the same parameters as [CLS08]. In their setting where costs are low and identical across items, PBDC is a special case of BSP. However, it still attains between 97.5% and 100% of the (nearly optimal) BSP profit. If we allow costs to vary and be more significant, then PBDC drastically outperforms other simple mechanisms (PC, PB, BSP), demonstrating its robustness under different scenarios. In fact, the worst case for PBDC is the aforementioned setting where it attains 97.5% of the profit of the best simple mechanism; contrast this with 79.9%, 16.8%, 59.5% for PC, PB, BSP, respectively, in their worst-case settings. In addition to being profit-maximizing, PBDC also achieves excellent global surplus in our simulations. Finally, we show that PBDC attains between 96.6% and 99.4% of the optimal profit (which prices all subsets) when $n = 3$, and scales well as n increases.

The general goal of our work is to dispel the myth that high costs should drive a firm away from bundling and toward individual sales. PBDC allows the firm to reap the benefits of bundling while preventing items from being consumed for utility below cost. We should point out that there do exist costless examples with independent valuations on which PBDC performs poorly relative to the optimal mechanism (which is why it is necessary to include PC in the statement of the second theoretical result). Here is a list, along with why PBDC (equivalent to PB) is ill-advised for each instance:

- Example 15 from [HN12]: there are various different valuations, each of which realizes to an exorbitant value with a small probability; bundling is ineffective because the probability that more than one valuation is non-zero is infinitesimal
- Examples 1 and 2 from [Rub16]: there is a need to *partition* the items, i.e. split them into groups, and offer each group as a different bundle
- Example 4.2 from Section 4 of our paper: there is a need to price-discriminate, i.e. offer high individual prices and a discounted bundle price

However, our numerical experiments demonstrate that over “average” instances, PBDC performs far better than these pathological constructions and the worst-case bound of $\frac{1}{5.2} \approx 19.2\%$ suggest.

Indeed, once PBDC has eliminated the effect of costs, selling everything under one bundle leaves very little to be desired.

1.1 Literature Review

Bundling has been the focus of many papers in three different disciplines: economics, computer science, and operations research/marketing science. In general, the literature can be classified into three categories: papers that provide insights, papers that suggest approximate algorithms with attractive worst-case bounds, and papers that develop computationally efficient algorithms. In this subsection we attempt to highlight the most important contributions to the bundling literature, independent of discipline.

Two Items. In the economics literature, the earliest recognition of bundling being able to increase the revenue from selling two items is usually attributed to [Sti63]; other early research for two products includes [AY76, Sch84, MMW89]. Since then, [VK03, MRT07] have established conditions for bundling being optimal for two potentially correlated goods.

Simple Mechanisms. For more than two items, there is a great practical interest in finding simple pricing schemes that are both profitable and easy to explain to the customer; for surveys on how bundling has affected marketing practice see [ST02, VM09]. However, the only concrete, general pricing scheme we have found in this literature, other than the classical PC and PB, is the BSP proposed by [HC05] and [CLS08]. Our scheme, PBDC, attempts to add to this literature by providing a transparent, easy-to-compute heuristic.

Most of the attempts to prove that simple pricing schemes are indeed capturing most of the optimal profit have been restricted to special cases (see [MV06, MV07]), or empirical evidence, as in the case of BSP, where its great experimental success has been unexplained. That’s where we turn to the computer science literature. There has been more general work on auctions with multiple buyers, or valuation functions where the valuation for a subset may not be additive over the items in the set, for which we refer to [RW15, Yao15] and the references therein. We focus on the case of a single buyer with additive valuations, which is the bundling problem.

In this special case, [HN12] introduce performance guarantees for simple mechanisms, which are further studied in [HR12, HN13]. One line of work ([LY13, BILW14]) culminated in a proof that either PB or PC must be within $\frac{1}{6}$ of optimal, for arbitrary independent valuations. By relating PBDC to PB, and improving upon their techniques, we are able to prove that either PBDC or PC must be within $\frac{1}{5.2}$ of optimal for the independent case with costs. When all costs are equal, PBDC is a special case of BSP, so our work provides the first theoretical explanation for some of the successes in [CLS08].

Recently, mechanisms where items are partitioned before bundling have also been advocated as simple in [CH13, Rub16]. Our bound improves the theoretical guarantee for the partitioning scheme in [Rub16] from $\frac{1}{6}$ to $\frac{1}{5.2}$. The same core-tail decomposition of [BILW14] has also been recently used in [BDHS15, Yao15], so our new bound improves guarantees from those works as well.

Computational Solutions. Others have tried to tackle the problem with more items by giving up on simplicity and computing an explicit optimal or near-optimal solution using optimization techniques. A mixed integer programming formulation was first seen in [HM90], and recently in the mechanism design literature, explicit polynomial-time solutions were provided via linear programming in [CDW12, CH13].

As far as computing the optimal prices for simple mechanisms, [WHCA08] use non-linear mixed integer programming to solve for the optimal BSP prices, while [Rub16] gives a PTAS for the optimal partitioning. Computation is another benefit of PBDC—like PB, it only requires calculating one price, which can be done via convolution.

Large Number of Items. Yet another line of work addresses the complexity of many items by claiming that PB is guaranteed to be optimal as the number of items approaches infinity, assuming independence and uniformly bounded variances. Traditionally, this line of work has dealt with information goods which have no marginal costs (see [BB99, BB00]), or showed that costs have a substantial effect on the efficacy of PB (see [FN06, IW10]). [Arm99] advocates that the same result can be achieved with costs by using a cost-based two-part tariff, which we prove is equivalent to PBDC.

Our research strengthens this line of work by using Cantelli’s one-sided concentration inequality to get a tighter problem-dependent bound. Furthermore, we advocate for PBDC even on a small number of items, both with our problem-independent bound, and our numerical experiments.

Closed-form Solutions. There is also interest in finding analytical closed-form solutions for the optimal pricing under simple cases of the problem. In the case of two independent valuations, one of which is uniform on $[0, b_1]$ and the other which is uniform on $[0, b_2]$, [Eck10] derives elementary equations for the optimal Mixed Bundling prices. [Bha13] shows that the equations involve roots of high-degree polynomials once costs are introduced, and uses a linear approximation to record solutions. Our transformation in Section 2 shows that the problem with costs is equivalent to the problem for distributions uniform on $[a_1, b_1]$ and $[a_2, b_2]$, where a_1 and a_2 could be negative. The difficulty of analytical solutions in general is discussed in [Wil93, Arm96, PVM10].

1.2 Summary of Contributions and Outline of Paper

- We introduce the idea of PBDC, eliminating the problem costs pose to bundling (Section 2):
 - We show that PBDC has equivalent formulations in the *cost-based two-part tariff* that exists in the theoretical literature, as well as a new *per-item discount* scheme
 - The idea of PBDC motivates a transformation from costs to negative valuations, enabling the analysis in subsequent sections
- We improve “large-number-of-items” bounds for the performance of PBDC, using Cantelli’s inequality (Section 3):
 - We recover existing bounds asymptotically and achieve a better bound when the coefficient of variation is large

- Our bound suggests that the firm should not price the bundle such that the profit margin is less than $1/3$ of the expected welfare
- We provide finite-item, distribution-free bounds for the performance of PBDC (Section 4):
 - We generalize existing bounds to the case with costs, where PBDC is needed instead of PB
 - We improve existing bounds in both directions (upper and lower bound)
 - We compare this type of performance guarantee to that in the previous section
- We provide a continuation of the numerical experiments from [CLS08], demonstrating the efficacy of PBDC for a finite number of items (Section 5)

2 Set-up and Equivalence Propositions

A firm has n different items for sale. For each i , the cost incurred by the firm for selling item i is c_i , a non-negative real number. c_i can be thought of as an instantaneous production cost, the opportunity cost of saving the inventory for someone else, or the value of the item to the seller.

Each of the firm's customers has a valuation vector $x \in \mathbb{R}^n$ for the items. A customer wants at most one of each item, and her utility for a subset of items S is $\sum_{i \in S} x_i$. x can be thought of as a random vector drawn from D , the distribution of valuation vectors across the population. The firm is risk-neutral and its objective is to maximize the expected per-customer profit.

In the full generality of the problem, the firm's mechanism for selling the items is a *menu* \mathcal{M} of *entries* (q, s) , where $q \in [0, 1]^n$ is the *allocation* indicating the fraction of each item transferred to the customer, and s is the *payment* that must be made for this allocation. If q has fractional entries, then the allocation can be thought of as a lottery where the customer only gets some items with a certain probability. The customer is also risk-neutral and chooses the entry maximizing her expected surplus, $q^T x - s$. We will assume that for every entry, the payment covers the expected cost for the firm to produce that allocation, i.e. $s \geq q^T c$, where $c = (c_1, \dots, c_n)$. Simultaneously removing all entries in the menu with $s < q^T c$ cannot decrease the profit, since this can only force a customer who previously selected an entry with negative profit to select an entry with non-negative profit.

Let \mathcal{X} denote the support of D . Given a menu \mathcal{M} , for all $x \in \mathcal{X}$, let $q_{\mathcal{M}}(x)$ denote the allocation chosen by a customer with valuation vector x , and $s_{\mathcal{M}}(x)$ denote the corresponding payment. We will omit the subscript \mathcal{M} when the context is clear. $(q_{\mathcal{M}}(x), s_{\mathcal{M}}(x))$ must maximize the surplus of the customer among all entries of \mathcal{M} ³ (the mechanism is *incentive-compatible*), and one of these entries must be the no-purchase option with $q = 0, s = 0$ (the mechanism is *individually rational*).

³In the case there are multiple maxima, the firm can choose between them; this can always be achieved by small perturbations.

The firm's profit maximization problem can be written as $\max_{\mathcal{M}} \mathbb{E}_{x \sim D}[s_{\mathcal{M}}(x) - q_{\mathcal{M}}(x)^T c]$. However, the optimization over general menus is intractable, and the resulting menu may not be practical.

Definition 2.1. A *pricing scheme* is a restricted class of menus, implied by a compact way of communicating the menu to the customer. It is assumed that the optimization problem over the restricted class of menus can be solved efficiently.

The following pricing schemes are frequently referenced throughout this paper:

1. Pure Components (PC): items are offered individually, at respective non-negative prices $P_1^{\text{PC}}, \dots, P_n^{\text{PC}}$.
2. Pure Components with Uniform Discount (PCUD): items are offered individually, at respective prices $P_1^{\text{PCUD}}, \dots, P_n^{\text{PCUD}}$. There is an absolute discount of P_0^{PCUD} which is applied to each item bought beyond the first. For all i , $P_i^{\text{PCUD}} \geq P_0^{\text{PCUD}} \geq 0$ is imposed.
3. Pure Bundling (PB): all of the items are offered in a single bundle at non-negative price P_0^{PB} , and there are no separate sales.
4. Pure Bundling with Disposal (PBD): all of the items are offered in a single bundle at price P_0^{PBD} , with the agreement that if the customer buys the bundle, she can return any subset of items S for a refund equal to $\sum_{i \in S} P_i^{\text{PBD}}$. For all i , $P_i^{\text{PBD}} \geq 0$ is imposed. Furthermore, $P_0^{\text{PBD}} \geq \sum_{i=1}^n P_i^{\text{PBD}}$ is imposed.
5. Tariff Pricing (TP): there is a membership fee of P_0^{TP} to enter the market. If the customer enters the market, she can buy up to one unit of each item i for the price of P_i^{TP} . $P_0^{\text{TP}}, P_1^{\text{TP}}, \dots, P_n^{\text{TP}}$ are all imposed to be non-negative.
6. Bundle-Size Pricing (BSP): the price of a subset S is $P_{|S|}^{\text{BSP}}$, which is dependent only on the number of items in S , and not which items are in S . $0 = P_0^{\text{BSP}} \leq P_1^{\text{BSP}} \leq \dots \leq P_n^{\text{BSP}}$ is imposed.

PC and PB were introduced by [AY76]. PCUD and PBD are variations of PC and PB, respectively, and to the best of our knowledge, PCUD and PBD are new to our paper. BSP was introduced by [HC05, CLS08], while the idea of TP could be seen in [Arm99]. Note that PC corresponds to the degenerate case of PBD where $P_0^{\text{PBD}} = \sum_{i=1}^n P_i^{\text{PBD}}$.

Our first observation is the following:

Proposition 2.2. *PCUD, PBD, and TP represent the same class of menus. Specifically, given a PBD representation with prices*

$$(P_0^{\text{PBD}}, P_1^{\text{PBD}}, \dots, P_n^{\text{PBD}}) \tag{1}$$

the corresponding PCUD representation is

$$(P_0^{\text{PCUD}} = P_0^{\text{PBD}} - \sum_{i=1}^n P_i^{\text{PBD}}, P_1^{\text{PCUD}} = P_1^{\text{PBD}} + P_0^{\text{PCUD}}, \dots, P_n^{\text{PCUD}} = P_n^{\text{PBD}} + P_0^{\text{PCUD}}) \tag{2}$$

and the corresponding TP representation is

$$(P_0^{\text{TP}} = P_0^{\text{PBD}} - \sum_{i=1}^n P_i^{\text{PBD}}, P_1^{\text{TP}} = P_1^{\text{PBD}}, \dots, P_n^{\text{TP}} = P_n^{\text{PBD}}). \quad (3)$$

The proofs of propositions are deferred to Appendix A. Hereinafter, we will always refer to PBD instead of PCUD and TP for the analysis; however, the existence of PCUD and TP gives the firm additional flexibility in how to describe these menus to the customer. Specifically, when the number of items is small, PCUD should be used instead of TP, since it does not sound so enticing for one to pay a surcharge in order to buy a small number of items. On the other hand, when the number of items is large, P_0^{PCUD} tends to be large, causing the individual items to be marked at exorbitant prices should PCUD be used. In this case, paying a membership fee to have access to all the items does not sound so bad.

We are especially interested in the following subclass of PBD, where $P_i^{\text{PBD}} = c_i$ for all i . Similar subclasses can also be defined for PCUD and TP, following Proposition 2.2.

Pure Bundling with Disposal for Cost (PBDC): the bundle with all of the items is offered at P_0^{PBD} . If the customer buys the bundle, she can return any subset S of items for a refund of $\sum_{i \in S} c_i$.

Setting the refund for each item equal to its cost is a logical restriction to put on PBD. To see why, consider the following definitions:

Definition 2.3. The *welfare* generated by a customer with valuations (x_1, \dots, x_n) is $\sum_{i=1}^n \max\{x_i - c_i, 0\}$, which is realized when every item valued above cost is transferred and no other items are transferred. Welfare can be split up as follows:

- The *total surplus* is the welfare realized from transfers that occurred, equal to $\sum_{i=1}^n q_i(x)(x_i - c_i)$. Total surplus can be further split up depending on the price charged:
 - The *producer surplus* is another term for the profit earned by the firm, equal to $s(x) - \sum_{i=1}^n q_i(x)c_i$.
 - The *consumer surplus* is the utility gained by the customer, equal to $\sum_{i=1}^n q_i(x)x_i - s(x)$.
- The *deadweight loss* is the welfare lost because an item valued above cost was not transferred, equal to $\sum_{i: x_i > c_i} (1 - q_i(x))(x_i - c_i)$.
- The *overinclusion loss*⁴ is the welfare lost because items were consumed for utility below cost, equal to $\sum_{i: x_i < c_i} q_i(x)(c_i - x_i)$.

It is clear from the equations that the sum of producer surplus, consumer surplus, deadweight loss, and overinclusion loss is $\sum_{i: x_i > c_i} (x_i - c_i)$, equal to welfare. Also, the fact that the consumer surplus is non-negative (since the customer can always choose the no-purchase option) implies that the profit is no greater than the total surplus, which in turn is no greater than the welfare.

⁴We use this terminology because [AY76] refer to the act of not incurring this loss as *exclusion*.

PBDC (and thus PBD) is strictly better than PB in the following sense:

Proposition 2.4. *Given a PB menu with price P^{PB} which is at least $c_1 + \dots + c_n$, consider instead the PBD menu with prices $(P_0^{\text{PBD}} = P^{\text{PB}}, P_1^{\text{PBD}} = c_1, \dots, P_n^{\text{PBD}} = c_n)$. For all x :*

- *The producer surplus is no less than before.*
- *The consumer surplus is no less than before.*
- *The deadweight loss is no more than before.*
- *The overinclusion loss is no more than before.*

Note that the preceding statements are not only in expectation; for *every* valuation vector x both the firm and the customer are better off. There is no reason to use PB if PBDC can be used instead, because PBDC is effectively PB where all valuations x_i have been replaced by $\max\{x_i, c_i\}$. This observation leads us to the following lemma:

Proposition 2.5. *The firm’s problem of maximizing expected profit with distribution D and costs c is equivalent to the transformed problem of maximizing expected revenue with distribution D' , where D' is the distribution D shifted downward by c_i in every dimension i . Furthermore, for any menu in the original problem, the corresponding menu in the transformed problem has the payment for each allocation reduced by the cost of producing that allocation.*

Proposition 2.5 is stated in precise mathematical language and proven in Appendix A. If the original optimization problem was over a restricted class of menus, then the class restriction in the transformed setting can be found via the second statement in Proposition 2.5.

For the remainder of this paper, we focus on bounding the revenue of PBDC in the transformed setting, which is more amenable to analysis. PBDC becomes the class of menus that offer the same price P for any non-empty subset of items (see Remark 7.1 in Appendix A for a technical proof of this). The customer makes a purchase if and only if her non-negative valuations (corresponding to valuations no less than cost) sum to at least P , in which case the firm earns P .

It may be tempting to truncate all negative customer valuations to 0 and claim that after this further transformation, PBDC is identical to PB. However, in Section 4, we bound the performance of the best PBDC menu relative to the *optimal* menu (with no restriction to a pricing scheme), which can be designed to exploit negative valuations to reduce the cannibalization of higher-priced menu entries. In general, revenue is *non-monotone*, i.e. increasing customer valuations can decrease the optimal revenue—see [HN12].

3 Asymptotic Performance Bounds

In this section we analyze the performance of PBDC with a large number of items, whose costs have been transformed into negative valuations according to the previous section. We assume that the valuations for different items are *independent* random variables. Also making some assumptions

on the means and variances of the individual distributions, PBDC is asymptotically optimal as the number of items becomes large.

[Arm99] has already proven this result for *Cost-based Tariff Pricing* (TP with the additional restriction that $P_i^{\text{TP}} = c_i$ for all i), which is equivalent to PBDC via Proposition 2.2. However, our analysis works under weaker assumptions, by employing Cantelli’s inequality, along with other tools. To our knowledge, we are the first to use Cantelli’s one-sided concentration inequality to get an improved performance bound for bundling; previous works by [BB99, Arm99, FN06] all use the weaker Chebyshev’s inequality. The analysis also motivates our finite-item, distribution-free bounds in Section 4, where we again make improvements using Cantelli’s inequality.

Lemma 3.1. (*Cantelli’s Inequality*) *Let X be a random variable with (finite) mean μ and variance σ^2 . Let t be an arbitrary non-negative real number. Then*

$$\mathbb{P}[X - \mu \leq -t] \leq \frac{\sigma^2}{\sigma^2 + t^2}$$

We refer the reader to [Lug09] for a proof, as well as more background. Our main result in this section is the following:

Theorem 3.2. *Suppose a firm is selling items to a customer with valuation vector x drawn from distribution D . Let $\text{VAL}^+(D)$ denote the mean of the welfare, equal to $\mathbb{E}_{x \sim D}[\sum_i \max\{x_i, 0\}]$, and assume that $0 < \text{VAL}^+(D) < \infty$. Furthermore, let $\text{Cv}^+(D)$ denote the coefficient of variation of the welfare, equal to $\frac{\sqrt{\text{Var}_{x \sim D}[\sum_i \max\{x_i, 0\}]}}{\mathbb{E}_{x \sim D}[\sum_i \max\{x_i, 0\}]}$, and assume that $\text{Cv}^+(D) < \infty$. Then for all $\varepsilon \in [0, 1]$, the expected revenue of the PBDC menu with price $(1 - \varepsilon)\text{VAL}^+(D)$ is at least $\frac{\varepsilon^2 - \varepsilon^3}{\varepsilon^2 + (\text{Cv}^+(D))^2} \cdot \text{VAL}^+(D)$. In particular, if*

$$\varepsilon = \frac{2(\text{Cv}^+(D))^{\frac{2}{3}}}{3(\text{Cv}^+(D))^{\frac{2}{3}} + 2}, \tag{4}$$

then the expected revenue is at least

$$\frac{4}{4 + 24(\text{Cv}^+(D))^{\frac{2}{3}} + 45(\text{Cv}^+(D))^{\frac{4}{3}} + 27(\text{Cv}^+(D))^2} \cdot \text{VAL}^+(D) \tag{5}$$

which in turn is at least

$$(1 - 6(\text{Cv}^+(D))^{\frac{2}{3}}) \cdot \text{VAL}^+(D). \tag{6}$$

(6) shows that when the coefficient of variation is close to 0, ε scales with $(\text{Cv}^+(D))^{\frac{2}{3}}$ and earns a $(1 - \Theta((\text{Cv}^+(D))^{\frac{2}{3}}))$ -fraction of the expected welfare, recovering the result from [BB99] and [Arm99]. However, for larger $\text{Cv}^+(D)$, we still get a non-zero revenue guarantee in (5), and interestingly our analysis never recommends offering the bundle below the price of $(1 - \frac{2}{3})\text{VAL}^+(D) = \frac{\text{VAL}^+(D)}{3}$. Contrast this to the previous analyses, which recommend $\varepsilon = 1$ when $\text{Cv}^+(D) \geq 1$, earning zero revenue. The value of ε in (4), recommended by our analysis, is useful even when the firm has the resources to compute the optimal value of ε from D —both as a managerial reference point, as well as in situations where the firm knows the mean and variance in demand but is uncertain about the exact distribution.

Theorem 3.2 treats the welfare as an abstract random variable, but the revenue guarantee is weak if the coefficient of variation is large. Independence is important in allowing the “law of large numbers” to control $\text{Cv}^+(D)$ when the number of items n is large.

Corollary 3.3. *Suppose a firm is selling n items to a customer with independent valuations x_1, \dots, x_n forming product distribution D . Let μ_{\min} be a lower bound on $\mathbb{E}[\max\{x_i, 0\}]$, and let σ_{\max}^2 be an upper bound on $\text{Var}[\max\{x_i, 0\}]$, over $i = 1, \dots, n$. Suppose $\mu_{\min} > 0$, $\sigma_{\max} < \infty$, and $n > (\frac{\sigma_{\max}}{\mu_{\min}})^2$. Then the expected revenue of an optimal menu within PBDC is at least*

$$(1 - 6(\frac{\sigma_{\max}}{\mu_{\min}})^{\frac{2}{3}} \frac{1}{\sqrt[3]{n}}) \cdot \text{VAL}^+(D).$$

Taking $n \rightarrow \infty$, we get the result that PBDC extracts the entire welfare. Note that truncating the random variables x_i from below by 0 can only increase the mean and decrease the variance, so any lower bound on $\mathbb{E}[x_i]$ and upper bound on $\text{Var}[x_i]$ would also satisfy the conditions in Corollary 3.3.

Proof. Proof of Theorem 3.2. Let $X = \sum_{i=1}^n \max\{x_i, 0\}$ be a single random variable representing the welfare of a valuation vector drawn from D . As additional shorthand, let $\mu = \text{VAL}^+(D)$ denote the mean of X , $\sigma = \text{VAL}^+(D) \cdot \text{Cv}^+(D)$ denote the standard deviation of X , and $C = \text{Cv}^+(D)$ denote the coefficient of variation of X .

We would like to bound the probability that $X < (1 - \varepsilon)\mu$ from above. Applying Cantelli’s inequality with $t = \varepsilon\mu$, this probability is at most $\frac{\sigma^2}{\sigma^2 + \varepsilon^2\mu^2}$. Therefore, our expected revenue is at least

$$(1 - \varepsilon)\mu \cdot (1 - \frac{\sigma^2}{\sigma^2 + \varepsilon^2\mu^2}) = \mu \cdot \frac{(1 - \varepsilon)\varepsilon^2\mu^2}{\sigma^2 + \varepsilon^2\mu^2}.$$

The fraction of expected welfare earned is

$$\frac{\varepsilon^2 - \varepsilon^3}{\varepsilon^2 + C^2} \geq \frac{\varepsilon^2 - \varepsilon^3}{\frac{2}{3}y^3C^{-\frac{2}{3}} + \frac{1}{3}C^{\frac{4}{3}} + C^2} \tag{7}$$

$$\geq \frac{\varepsilon^2 - (1 + \frac{2}{3}C^{-\frac{2}{3}})\varepsilon^3}{\frac{1}{3}C^{\frac{4}{3}} + C^2}. \tag{8}$$

The first inequality uses the *weighted arithmetic mean–geometric mean inequality* (see [Zha08] for a reference), which yields $\frac{2y^3+C^2}{3} \geq (y^6C^2)^{\frac{1}{3}} = y^2C^{\frac{2}{3}}$. The second inequality is because for a fraction $\frac{a}{b}$ with $0 < a \leq b$, subtracting the same amount less than b from both the numerator and the denominator can only decrease the fraction.

Now, if we choose $\varepsilon = \frac{2C^{\frac{2}{3}}}{3C^{\frac{2}{3}} + 2}$ (this is motivated by setting the derivative of (8) to zero), then

the LHS of (7) becomes

$$\begin{aligned}
\frac{4C^{\frac{4}{3}}(1 - \frac{2}{3})}{(3C^{\frac{2}{3}} + 2)^2(\frac{1}{3}C^{\frac{4}{3}} + C^2)} &= \frac{\frac{4}{3}}{(2 + 3C^{\frac{2}{3}})^2(\frac{1}{3} + C^{\frac{2}{3}})} \\
&= \frac{4}{4 + 24C^{\frac{2}{3}} + 45C^{\frac{4}{3}} + 27C^2} \\
&= 1 - 6C^{\frac{2}{3}} \left(\frac{4 + \frac{15}{2}C^{\frac{2}{3}} + \frac{9}{2}C^{\frac{4}{3}}}{4 + 24C^{\frac{2}{3}} + 45C^{\frac{4}{3}} + 27C^2} \right) \\
&\geq 1 - 6C^{\frac{2}{3}}
\end{aligned}$$

where the inequality holds because the expression in parentheses is less than 1. This establishes both (5) and (6), completing the proof of Theorem 3.2. \square

Proof. Proof of Corollary 3.3. By independence, $\text{Var}[\sum_{i=1}^n \max\{x_i, 0\}] = \sum_{i=1}^n \text{Var}[\max\{x_i, 0\}]$ which is at most $n\sigma_{\max}^2$. Furthermore, $\mathbb{E}[\sum_{i=1}^n \max\{x_i, 0\}] \geq n\mu_{\min}$. Therefore, $\text{Cv}^+(D)$ is upper bounded by $\frac{\sigma_{\max}}{\mu_{\min}\sqrt{n}}$, and it is easy to see from the proof of Theorem 3.2 that all of its statements continue to hold when $\text{Cv}^+(D)$ is replaced by an upper bound on $\text{Cv}^+(D)$. The condition $n > (\frac{\sigma_{\max}}{\mu_{\min}})^2$ ensures that $\text{Cv}^+(D) < 1$, and the result follows immediately from substituting $\text{Cv}^+(D) \leq \frac{\sigma_{\max}}{\mu_{\min}\sqrt{n}}$ into (6). \square

4 Finite-item, Distribution-free Performance Bounds

In this section we analyze the performance of PBDC with only the independence assumption on the items, whose costs have been transformed into negative valuations according to Section 2. All proofs are deferred to Appendices B–C, but we sketch the techniques needed to handle arbitrary distributions.

Theorem 4.1. *Suppose a firm is selling items to a customer with independent (and potentially negative) valuations forming product distribution D . Let $\text{REV}(D)$ denote the expected revenue of an optimal menu (along with tie-breaking rules) for distribution D . Then the expected revenue of either the optimal menu within PBDC or the optimal menu within PC is at least*

$$\frac{1}{5.2} \cdot \text{REV}(D).$$

In the previous section, we showed that with assumptions on the number of items and their variances, PBDC can earn almost all of the expected welfare, $\text{VAL}^+(D)$. However, this is clearly false without distributional assumptions— $\text{VAL}^+(D)$ can be infinite. To recover some guarantee on performance, we need to use the *core-tail decomposition*, a technique developed through [HN12, LY13, BILW14].

The idea of the core-tail decomposition is to split off from each independent distribution all the valuations above a large positive cutoff (the “tail”). The remaining valuations (the “core”) are bounded, and it can be shown using a concentration inequality that PB (in our case PBDC)

Table 1: Comparison of Guarantees

	Corollary 3.3 (Section 3)	Theorem 4.1 (Section 4)
Dependence on n	bound only relevant for large n	none
Assumptions on Distributions	uniformly bounded variance	none
Benchmark	expected welfare	expected revenue of optimal menu
Fraction of Benchmark Guaranteed	100% as $n \rightarrow \infty$	$\frac{1}{5.2}$, with the help of PC
Related Literature	[BB99] [Arm99]	[BILW14]

*The advantages of each bound are bolded.

performs well relative to the welfare of the core. Meanwhile, PC can be shown to perform well relative to the *optimal mechanism* in the tail. Finally, the core bound (relative to the expected welfare of the core) and the tail bound (relative to the optimal expected revenue for the tail) can be combined to get a performance guarantee relative to the optimal expected revenue on D .

Theorem 4.1 improves upon the main result of [BILW14] by increasing the guarantee from $\frac{1}{6}$ to $\frac{1}{5.2}$, and allowing for negative valuations. The differences in our analysis can be summarized as follows:

- We analyze the core and tail together, and show that the worst case for PBDC in the core and worst case for PC in the tail cannot simultaneously occur
- We use Cantelli’s inequality instead of Chebyshev’s inequality in the core bound
- We show that the core bound and the tail bound can still be combined to upper-bound the optimal revenue on D when the optimal mechanism can exploit negative valuations

Table 1 compares Theorem 4.1 to the type of bound in the previous section, in particular Corollary 3.3. Essentially, to accommodate arbitrary distributions, we have to settle for a constant fraction of the optimum, compare against an optimum that is convoluted, and also allow ourselves to use PC in pathological cases.

One additional point worth mentioning is that it is unclear from Theorem 4.1 what the optimal prices for PBDC or PC are. It is assumed that the firm, knowing distribution D , can compute the optimal prices for both PBDC and PC and use the scheme with higher expected revenue, with the knowledge that it will be within $\frac{1}{5.2}$ of optimal. Meanwhile, Theorem 3.2, with its simpler analysis, has an explicit benchmark price of $\frac{(Cv^+(D))^{2/3}+2}{3(Cv^+(D))^{2/3}+2} \cdot \text{VAL}^+(D)$ for the bundle in PBDC.

Finally, we address the tightness of Theorem 4.1. First we present a theoretical upper bound.

Example 4.2. Consider an instance with 2 costless items, which have IID valuations distributed as follows. There is a point mass of size $1 - \rho$ at 0, a point mass of size $\frac{\rho}{2}$ at 2, and the remaining $\frac{\rho}{2}$ mass distributed in an *equal-revenue* fashion on $[1, 2)$, i.e. selling individually at any price in $[1, 2)$

results in the same revenue. Formally, if Y is a random variable with this distribution, then

$$\mathbb{P}[Y \geq y] = \begin{cases} 1 & y = 0 \\ \rho & 0 < y \leq 1 \\ \frac{\rho}{y} & 1 \leq y \leq 2 \end{cases}$$

where the value of ρ is optimized to be $\frac{3}{3+\ln 2} \approx 0.81$.

Theorem 4.3. *Consider the instance in Example 4.2. The best possible PC revenue is 2ρ , attained by selling individual items at any price in $[1, 2]$. The best possible PB revenue is also 2ρ , attained by selling the bundle at the price of 2 or 3. The optimal revenue is at least $2\rho(2 - \rho)$; this value can be achieved by selling individual items at the price of 2, and the bundle at the discounted price of 3.*

Therefore, neither PC nor PB can obtain more than $\frac{3+\ln 2}{3+2\ln 2} \cdot \text{REV}(D)$ which is approximately

$$\frac{1}{1.19} \cdot \text{REV}(D).$$

In Example 4.2, both PC and PB perform poorly because there is a need to *price-discriminate*, i.e. allow customers who highly value an item to buy it for its individual price, but still give customers with lower valuations a chance of buying it as part of a discounted bundle. Very recently, [Rub16] constructed an example where both PC and PB perform poorly because there is a need to *partition* the items, i.e. split them into groups, and offer each group as a different bundle. In his example, the better of PC and PB can only obtain $\frac{1}{2} + \epsilon$ of the revenue via partitioning, which is smaller than our bound. However, our example exhibits the worst-known loss from not price-discriminating, where even partitioning performs poorly relative to the optimal mechanism. Our example also only requires two IID items, following the examples of [HN12, HR12]; the example in [Rub16] requires a large number of distinct items.

Nonetheless, there is a large gap between the best-known lower bound from Theorem 4.1 and the best-known upper bounds, and furthermore, being guaranteed only $\frac{1}{5.2} \approx 19.2\%$ of the optimal profit does not sound so enticing. However, this bound arises from a worst-case analysis that needs to address pathological instances, on which PBDC does not obtain $\frac{1}{5.2}$ of the optimum, but PC does. In the next section, we test the performance of PBDC over “average” instances.

5 Numerical Experiments

In this section we conduct a continuation of the numerical experiments from [CLS08] where PBDC is included as an additional pricing scheme. As a disclaimer, we should quote [CLS08] on the limitations inherent to this kind of numerical analysis:

“Although we attempt to cover a large space of parameter values, the results clearly depend on the specific parameters we choose (i.e., the choice of grid). Further, there is no way for us to know whether we are under- or oversampling the relevant (i.e.,

Table 2: Ranges of Parameters, replicated from [CLS08]

Exponential	Marginal distributions are Exponential, with means chosen uniformly from $[0.2, 2]$. Thus the rates λ are in $[0.5, 5]$.
Logit	Marginal distributions are Gumbel, with fixed scale $\sigma = 0.25$ and means chosen uniformly from $[0, 2.5]$. Thus the locations μ are in $[-0.25\gamma, 2.5 - 0.25\gamma] \approx [-0.14, 2.36]$.
Lognormal	Marginal distributions are Lognormal. Logarithms of valuations are Normally distributed with means chosen uniformly from $[-1.5, 1]$ and fixed variance $\sigma^2 = 0.25$. Thus the original valuations have means in $[e^{-1.5+0.125}, e^{1+0.125}] \approx [0.25, 3.08]$.
Normal	Marginal distributions are Normal with means chosen uniformly from $[-1, 2.5]$ and variances chosen uniformly from $[0.25, 1.75]$.
Uniform	Marginal distributions are Uniform on $[0, b]$, where b is chosen uniformly from $[0.4, 4]$. Thus the means are in $[0.2, 2]$.

*Note that [CLS08] have two separate families of Normal distributions, one with varying mean and one with varying variance. For convenience, we allow both to vary at the same time.

empirically plausible) combinations of parameters. So, for example, when we describe average outcomes, these should certainly not be interpreted as outcomes that would be expected in an empirical sense they should be interpreted narrowly as the average of the experiments we performed.”

5.1 Procedure

For consistency, we follow the setup from [CLS08] as closely as possible. We use the same five families of valuation distributions commonly used to model demand—Exponential, Logit, Lognormal, Normal, and Uniform. We also use the same ranges of parameters for these families, as outlined in Table 2. The parameters were calibrated so that valuations across different families have similar means on average, and the highest means are around 10 times the lowest means. We allow for free disposal, just like [CLS08]—all negative valuations are converted to 0. We assume that valuations are independent across items.

As far as costs, we consider three scenarios:

1. *Heterogeneous Items*: valuation distributions fluctuate in accordance with Table 2, while costs are low. The cost of each item is set to 0.2, except in the case of Uniform distributions, where it is set to half the item’s mean valuation. These are the same numbers used in [CLS08], so this scenario is a duplicate of some of their experiments.
2. *Heterogeneous Costs*: valuation distributions are identical, while costs fluctuate. The costs are chosen uniformly from $[0, 2.5]$, approximately the same range as the means. In the case of the bounded Uniform distribution, the costs are chosen uniformly from 0 to 0.75 times the

Table 3: Summary of Parameters and Costs

Taste Distribution	Range for Means	Fixed Mean	Range for Costs	Fixed Cost
Exponential	[0.20,2.00]	1.25	[0,2.5]	0.2
Logit	[0.00,2.50]	1.5	[0,2.5]	0.2
Lognormal	[0.25,3.08]	$e^{0.5+0.125} \approx 1.87$	[0,2.5]	0.2
Normal	[0.00,2.50]	1.5 (and fixed variance 1)	[0,2.5]	0.2
Uniform	[0.20,2.00]	irrelevant	$[0, 1.5] \times \text{mean}$	$0.5 \times \text{mean}$

maximum valuation, so that there always are some customers who value the item above cost. The fixed valuation distributions are disclosed in Table 3—we choose a mean that is on the high end of the range to avoid degenerate instances, where the welfare in the system is near 0 when costs are high.

3. *Heterogeneous Items and Costs*: both valuation distributions and costs are allowed to fluctuate (independently) according to the preceding scenarios.

The parameters and costs are summarized in Table 3.

We compare the four simple pricing schemes—PC, PB, BSP, and PBDC. Unlike [CLS08], we do not compute the optimal deterministic profit with $2^n - 1$ prices, since it is hard to compute, difficult to implement in practice, and could be far off from the optimal profit of a randomized mechanism anyway. Skipping this expensive computation allows us to consider n from 2 up to 6.

For each combination of the 3 cost scenarios, 5 demand distributions, and 5 options for n , we randomly generate 200 instances, resulting in 15000 total instances. [CLS08] were able to discretize the parameter space for each combination and generate 220 instances in a grid. While generating instances in a grid is more reliable than generating instances randomly, we simply have too many combinations, because we allow costs to vary independently, allow for larger n , and in the case of Normal distributions, also allow variances to vary independently. Our randomized approach has the advantage of being scalable, and not depending on the exact grid chosen. Furthermore, we have verified that 200 instances per combination is enough, in that repeating the experiments does not cause the reported observations to change by any significance.

5.2 Observations

First, we report the performance of the simple pricing schemes separated by scenario. For each instance (out of the 15000), we compute which of PC, PB, BSP, PBDC earns the most profit on that instance, and record the performance of every pricing scheme as a *fraction* of this optimum. For each scenario (out of the 3), we report the median performance as well as 10'th percentile performance of every pricing scheme across the 1000 instances of each distribution family (200 for each of $n = 2, \dots, 6$), in Table 4. We also count the number of instances on which each pricing scheme was best, in Table 5.

Table 4: Median and 10'th Percentile Performances of Pricing Schemes

		Heterogeneous Items				Heterogeneous Costs				Both Heterogeneous			
		PC	PB	BSP	PBDC	PC	PB	BSP	PBDC	PC	PB	BSP	PBDC
Exponential	0.1 %ile	.766	.940	1	.994	.850	.269	.807	.995	.884	.137	.759	.978
	0.5 %ile	.835	.972	1	.999	.931	.489	.907	1	.964	.403	.926	1
Logit	0.1 %ile	.826	.937	1	.988	.815	.063	.245	.996	.852	.001	.385	.987
	0.5 %ile	.873	.992	1	.998	.891	.481	<i>.595</i>	1	.938	<i>.168</i>	.894	1
Lognormal	0.1 %ile	.734	.982	1	.998	.775	.513	.760	1	.852	.015	.327	.931
	0.5 %ile	<i>.799</i>	.996	1	1	.861	.730	.880	1	.970	.245	.887	1
Normal	0.1 %ile	.825	.745	1	.957	.858	.297	.779	.982	.904	.010	.699	.974
	0.5 %ile	.890	.880	1	<i>.975</i>	.926	.547	.912	1	.978	.198	.933	1
Uniform	0.1 %ile	.904	.834	.940	.949	.872	.348	.875	.948	.914	.380	.605	.937
	0.5 %ile	.959	.867	.975	.998	.933	.578	.974	1	.982	.638	.875	1

*For each scenario, the best performance in each row is **bolded**. The overall worst median performance of each pricing scheme is *italicized*.

We know from [CLS08] that BSP is within 1% of the deterministic optimum in most of their settings, so there is minimal room for improvement under scenario 1. In fact, PBDC is a special case of BSP when all costs are identical, and very similar to PB when costs are low. However, as one can see in Table 4, PBDC still extracts close to 100% of the BSP profit under this scenario, hence it also extracts close to 100% of the deterministic optimum. For Uniform valuations, PBDC is no longer a special case of BSP, since costs vary proportionally with means. PBDC actually outperforms BSP in this setting—indeed, this is by far the worst setting for BSP listed in [CLS08, tbl. 5], where it only extracts 91% of the deterministic optimum.

Scenario 2, where valuation distributions are identical but costs are allowed to fluctuate, really exhibits the power of PBDC, which allows customers to consume only the items they value above cost via self-selection. PC loses out on not bundling similar items that differ only in cost, while BSP is forced to compromise between charging cheap prices where high-cost items may be consumed for utility below cost, or charging expensive prices that result in a lot of deadweight loss in the low-cost items. In Appendix D, we show an instance that exemplifies why BSP performs so poorly when the costs in the setup from [CLS08] are increased.

When both valuation distributions and costs are allowed to vary under scenario 3, PBDC is still the best strategy by a significant margin. However, the benefits of bundling have decreased when items can be drastically different, so PC has gained ground. It seems intuitive to hypothesize that the performance of PC is inflated by the small values of n we are using. In the next subsection, we organize our reports separated by n , under scenario 3 (where both valuation distributions and costs are allowed to fluctuate).

Table 5: Number of Instances on which each Pricing Scheme was Best

	Heterogeneous Items				Heterogeneous Costs				Both Heterogeneous			
	PC	PB	BSP	PBDC	PC	PB	BSP	PBDC	PC	PB	BSP	PBDC
Exponential	5	-	995	-	19	-	114	867	206	-	113	681
Logit	0	-	1000	-	121	-	45	834	179	-	133	688
Lognormal	0	-	1000	-	8	-	70	922	245	-	68	687
Normal	12	-	988	-	20	-	180	800	216	-	201	583
Uniform	228	-	293	479	145	-	348	507	370	-	129	501

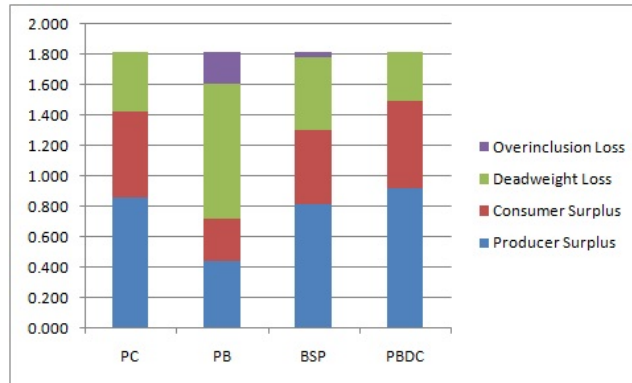


Figure 1: Breakdown of Welfare for each Pricing Scheme, averaged over n

5.3 Separation by n and Effects on Welfare

In this subsection, we allow both valuation distributions and costs to vary, and report averages across demand distributions, separated by n (instead of medians over the different choices for n , separated by demand distribution). Since the distribution families we’re amalgamating were calibrated to have similar means over their ranges of parameters, it makes sense in this subsection to report average absolute profits, instead of median fractions. We also report the figures defined in Definition 2.3, in the same way as [CLS08].

In Table 6, we report the expected values of these figures across the 1000 instances for each n . The main conclusions are best summarized in Figures 1-2.

The first graph (Figure 1) shows that although PBDC optimizes from the perspective of a selfish monopolist interested only in Producer Surplus, it has a similar advantage in Total Surplus. There is no Overinclusion Loss, and the monopolist is encouraged to choose a low tariff price so that most customers can enter the market. PC also incurs no Overinclusion Loss, but incurs more Deadweight Loss because it does not bundle. PB incurs significantly more Overinclusion Loss than any other strategy, forcing the customer into buying every item at once. All in all, PBDC is equally attractive from the standpoint of an altruistic policymaker interested in maximizing Total Surplus.

Table 6: Report of Economics Figures, separated by n

Number of Items	Statistic	PC	PB	BSP	PBDC
2	Producer Surplus	0.427	0.301	0.412	0.432
	Consumer Surplus	0.287	0.194	0.250	0.292
	Total Surplus	0.714	0.495	0.662	0.724
	Deadweight Loss	0.192	0.351	0.224	0.183
	Overinclusion Loss	-	0.061	0.021	-
3	Producer Surplus	0.655	0.395	0.630	0.683
	Consumer Surplus	0.437	0.254	0.382	0.436
	Total Surplus	1.092	0.649	1.011	1.119
	Deadweight Loss	0.291	0.604	0.352	0.264
	Overinclusion Loss	-	0.130	0.020	-
4	Producer Surplus	0.870	0.457	0.827	0.929
	Consumer Surplus	0.587	0.293	0.497	0.582
	Total Surplus	1.456	0.749	1.324	1.511
	Deadweight Loss	0.396	0.905	0.498	0.342
	Overinclusion Loss	-	0.198	0.031	-
5	Producer Surplus	1.070	0.504	1.030	1.167
	Consumer Surplus	0.705	0.297	0.595	0.703
	Total Surplus	1.775	0.802	1.625	1.870
	Deadweight Loss	0.488	1.158	0.600	0.394
	Overinclusion Loss	-	0.304	0.039	-
6	Producer Surplus	1.265	0.553	1.206	1.409
	Consumer Surplus	0.844	0.346	0.697	0.828
	Total Surplus	2.108	0.899	1.902	2.237
	Deadweight Loss	0.587	1.440	0.736	0.459
	Overinclusion Loss	-	0.356	0.057	-

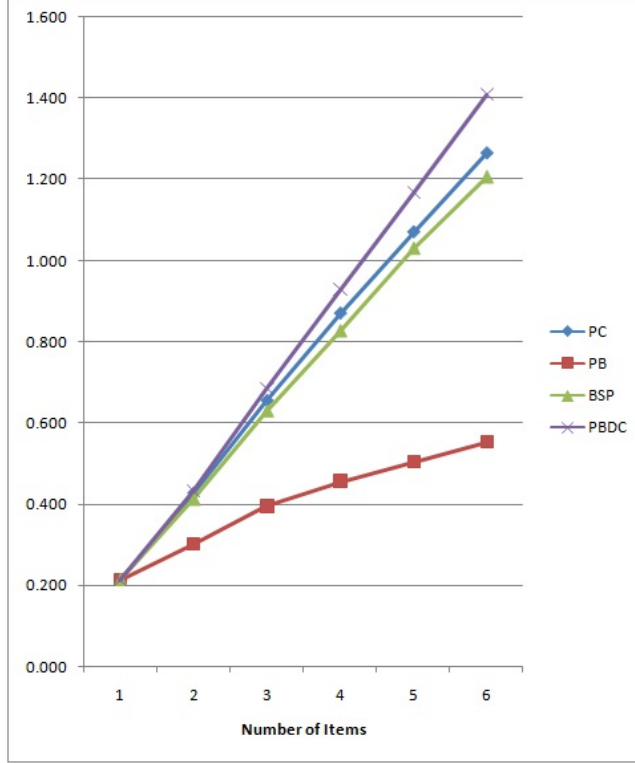


Figure 2: Average Profit of each Pricing Scheme, as a function of n

The second graph (Figure 2) shows the profits of each pricing scheme as n increases. The PC profits increase linearly with n , since items are sold separately. Both the PB and the BSP profits are concave in n —that is, the marginal gain from having one more item to sell is decreasing. Indeed, PB is burdened with adding to its grand bundle another item that could be valued below cost, while BSP is burdened with an additional distinct item to consider in its item-symmetric cost structure. PBDC is the only pricing scheme where the profit is convex in n , as each item creates additional incentive for the customer to enter the market, and makes their total utility from entering the market more concentrated. This confirms the hypothesis that while Table 4 reports a small gap between PC and PBDC under scenario 3, this gap quickly widens as n increases.

5.4 Grid Instances and Comparing with the Deterministic Optimum for $n = 3$

In this subsection, we generate instances in a grid where both valuation distributions and costs are allowed to vary, for the $n = 3$ case. There are 3 possibilities for distribution mean and 3 possibilities for cost for each of 3 different items, resulting in a total of $3^6 = 729$ instances. This is repeated over the 5 different demand distributions. The grid is outlined in Table 7; we centered the grid around the values from Table 3.

We report the performance of each simple pricing scheme over these 729 instances in the same manner as Table 4, except this time every number is recorded as a fraction of the optimal deter-

Table 7: Grid for Items Parameters and Costs

Taste Distribution	Grid for Means	Grid for Costs
Exponential	$\{0.5, 1.25, 2\}$	$\{0, 1.25, 2.5\}$
Logit	$\{0.5, 1.5, 2.5\}$	$\{0, 1.25, 2.5\}$
Lognormal	$\{e^{0.125}, e^{0.625}, e^{1.125}\} \approx \{1.13, 1.87, 3.08\}$	$\{0, 1.25, 2.5\}$
Normal	$\{0.5, 1.5, 2.5\}$ (and fixed variance 1)	$\{0, 1.25, 2.5\}$
Uniform	$\{0.4, 1, 1.6\}$	$\{0, 0.75, 1.5\} \times \text{mean}$

Table 8: Median and 10'th Percentile Performances over the Grid

Taste Distribution	Statistic	PC	PB	BSP	PBDC
Exponential	0.1 %ile	.863	.159	.600	.889
	0.5 %ile	.932	.474	.872	.966
Logit	0.1 %ile	.881	.000	.219	.966
	0.5 %ile	.941	.308	.918	.994
Lognormal	0.1 %ile	.834	.575	.735	.946
	0.5 %ile	.909	.898	.959	.989
Normal	0.1 %ile	.877	.095	.593	.944
	0.5 %ile	.925	.479	.880	.968
Uniform	0.1 %ile	.874	.340	.461	.899
	0.5 %ile	.922	.723	.888	.972

ministic profit, which is at least the profit of any simple pricing scheme. The results are displayed in Table 8.

In the median case, PBDC obtains between 96.6% to 99.4% of the deterministic optimum across the different demand distributions. This confirms both that PBDC is performing well relative to the optimal deterministic profit and not just other simple mechanisms, and that our earlier numbers with random instances are consistent.

To summarize our numerical experiments, we considered both scenarios with low costs and scenarios with high costs, and reported median performances over $n = 2, \dots, 6$ for different demand distributions. When costs are low, PC can earn as little as 79.9% of the profit of the optimal simple mechanism. When costs are high, PB can earn as little as 16.8% of the profit of the optimal simple mechanism, BSP can earn as little as 59.5%, and PC also falls behind as n increases. PBDC has the highest percentages overall, and is by far the most robust over different cost scenarios, always obtaining at least 97.5% of the profit of the optimal simple mechanism. We should point out that throughout our simulations, PBDC was also computationally much faster than BSP, requiring an optimization over 1 price instead of n .

6 Conclusion and Open Questions

In this paper, we propose a simple strategy for the multi-product pricing problem: Pure Bundling with Disposal for Cost, or PBDC. We prove that PBDC is asymptotically optimal. When there are only a small number of items, we still guarantee that either PBDC or PC earns at least $\frac{1}{5.2} \approx 19.2\%$ of the optimal profit, and our simulations suggest that this is closer to 96.6%-99.4% in the average case, and that PC is not needed. While this is worse than the 99% achieved by [CLS08] for BSP in their experiments with lower costs, the pricing problem becomes much harder when costs are significant, and the existing simple pricing schemes (including BSP) fall behind PBDC by a great deal. Yet, production costs exceeding mean valuations is a common occurrence in industry, where only a small fraction of a company's customers may have interest in any particular item.

One caveat with PBDC is that the prices do reveal production costs to the customer. If this is undesired, a potential remedy is optimizing prices over the larger class of Tariff Pricing (TP) strategies, which has $n + 1$ degrees of freedom and is guaranteed to be at least as profitable as PBDC. We believe that using TP instead of PBDC is very reasonable in practice, so long as the firm can accept the significant increase in computation time and decrease in the manager's ability to interpret the pricing.

However, the true demand distribution is never known, and must be constructed from data. When the given demand is prone to error, we hypothesize that there is additional benefit in choosing strategies that optimize one price at a time (such as PC, PB, PBDC) over strategies that optimize $\Theta(n)$ prices together (such as BSP, TP, MB). Besides, the theoretical guarantee for PBDC is no worse than that for TP, and PBDC is optimal as the number of items approaches infinity. We find it particularly interesting that as n increases and there are *more* potential prices to optimize, the benefit of optimizing only *one* price is greater.

All in all, PBDC captures the concentration effects of bundling and the selection effects of individual sales in a single heuristic that is computationally minimal and highly marketable. We hope our work on PBDC will have an impact on both the theory and practice of bundling, and be viewed as an effort to tie together the streams of research from three different disciplines: economics, computer science, and operations research.

References

- [Arm96] Mark Armstrong, *Multiproduct nonlinear pricing*, Econometrica: Journal of the Econometric Society (1996), 51–75.
- [Arm99] ———, *Price discrimination by a many-product firm*, The Review of Economic Studies **66** (1999), no. 1, 151–168.
- [AY76] William James Adams and Janet L Yellen, *Commodity bundling and the burden of monopoly*, The quarterly journal of economics (1976), 475–498.

- [BB99] Yannis Bakos and Erik Brynjolfsson, *Bundling information goods: Pricing, profits, and efficiency*, Management science **45** (1999), no. 12, 1613–1630.
- [BB00] ———, *Bundling and competition on the internet*, Marketing science **19** (2000), no. 1, 63–82.
- [BDHS15] MohammadHossein Bateni, Sina Dehghani, MohammadTaghi Hajiaghayi, and Saeed Seddighin, *Revenue maximization for selling multiple correlated items*, Algorithms-ESA 2015, Springer, 2015, pp. 95–105.
- [Bha13] Hemant K Bhargava, *Mixed bundling of two independently valued goods*, Management Science **59** (2013), no. 9, 2170–2185.
- [BILW14] Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S Matthew Weinberg, *A simple and approximately optimal mechanism for an additive buyer*, Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on, IEEE, 2014, pp. 21–30.
- [CDW12] Yang Cai, Constantinos Daskalakis, and S Matthew Weinberg, *An algorithmic characterization of multi-dimensional mechanisms*, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, ACM, 2012, pp. 459–478.
- [CH13] Yang Cai and Zhiyi Huang, *Simple and nearly optimal multi-item auctions*, Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2013, pp. 564–577.
- [CLS08] Chenghuan Sean Chu, Phillip Leslie, and Alan Sorensen, *Nearly optimal pricing for multiproduct firms*, Tech. report, National Bureau of Economic Research, 2008.
- [CLS11] ———, *Bundle-size pricing as an approximation to mixed bundling*, The American Economic Review (2011), 263–303.
- [Eck10] John C Eckalbar, *Closed-form solutions to bundling problems*, Journal of Economics & Management Strategy **19** (2010), no. 2, 513–544.
- [FN06] Hanming Fang and Peter Norman, *To bundle or not to bundle*, RAND Journal of Economics (2006), 946–963.
- [HC05] Lorin M Hitt and Pei-yu Chen, *Bundling with customer self-selection: A simple approach to bundling low-marginal-cost goods*, Management Science **51** (2005), no. 10, 1481–1493.
- [HM90] Ward Hanson and R Kipp Martin, *Optimal bundle pricing*, Management Science **36** (1990), no. 2, 155–174.
- [HN12] Sergiu Hart and Noam Nisan, *Approximate revenue maximization with multiple items*, Proceedings of the 13th ACM Conference on Electronic Commerce, ACM, 2012, pp. 656–656.

- [HN13] ———, *The menu-size complexity of auctions*, Proceedings of the fourteenth ACM conference on Electronic commerce, ACM, 2013, pp. 565–566.
- [HR12] Sergiu Hart and Philip J Reny, *Maximal revenue with multiple goods: Nonmonotonicity and other observations*, Center for the Study of Rationality, 2012.
- [IW10] Rustam Ibragimov and Johan Walden, *Optimal bundling strategies under heavy-tailed valuations*, Management Science **56** (2010), no. 11, 1963–1976.
- [LFCK13] Minqiang Li, Haiyang Feng, Fuzan Chen, and Jisong Kou, *Numerical investigation on mixed bundling and pricing of information products*, International Journal of Production Economics **144** (2013), no. 2, 560–571.
- [Lug09] Gábor Lugosi, *Concentration-of-measure inequalities*, <http://www.econ.upf.edu/~lugosi/anu.pdf>, 2009.
- [LY13] Xinye Li and Andrew Chi-Chih Yao, *On revenue maximization for selling multiple independently distributed items*, Proceedings of the National Academy of Sciences **110** (2013), no. 28, 11232–11237.
- [MMW89] R Preston McAfee, John McMillan, and Michael D Whinston, *Multiproduct monopoly, commodity bundling, and correlation of values*, The Quarterly Journal of Economics (1989), 371–383.
- [MRT07] Kevin F McCardle, Kumar Rajaram, and Christopher S Tang, *Bundling retail products: Models and analysis*, European Journal of Operational Research **177** (2007), no. 2, 1197–1217.
- [MV06] Alejandro M Manelli and Daniel R Vincent, *Bundling as an optimal selling mechanism for a multiple-good monopolist*, Journal of Economic Theory **127** (2006), no. 1, 1–35.
- [MV07] ———, *Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly*, Journal of Economic Theory **137** (2007), no. 1, 153–185.
- [Mye81] Roger B Myerson, *Optimal auction design*, Mathematics of operations research **6** (1981), no. 1, 58–73.
- [PVM10] Ashutosh Prasad, R Venkatesh, and Vijay Mahajan, *Optimal bundling of technological products with network externality*, Management Science **56** (2010), no. 12, 2224–2236.
- [Rub16] Aviad Rubinstein, *On the computational complexity of optimal simple mechanisms*, Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, ACM, 2016, pp. 21–28.
- [RW15] Aviad Rubinstein and S Matthew Weinberg, *Simple mechanisms for a subadditive buyer and applications to revenue monotonicity*, Proceedings of the sixteenth ACM conference on Electronic commerce, ACM, 2015, pp. 377–394.

- [Sch84] Richard Schmalensee, *Gaussian demand and commodity bundling*, Journal of business (1984), S211–S230.
- [ST02] Stefan Stremersch and Gerard J Tellis, *Strategic bundling of products and prices: A new synthesis for marketing*, Journal of Marketing **66** (2002), no. 1, 55–72.
- [Sti63] George J Stigler, *United states v. loew’s inc.: A note on block-booking*, Sup. Ct. Rev. (1963), 152.
- [VK03] R Venkatesh and Wagner Kamakura, *Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products**, The Journal of business **76** (2003), no. 2, 211–231.
- [VM09] R Venkatesh and Vijay Mahajan, *11 the design and pricing of bundles: a review of normative guidelines and practical approaches*, Handbook of pricing research in marketing (2009), 232.
- [WHCA08] Shin-yi Wu, Lorin M Hitt, Pei-yu Chen, and G Anandalingam, *Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach*, Management Science **54** (2008), no. 3, 608–622.
- [Wil93] Robert B Wilson, *Nonlinear pricing*, Oxford University Press, 1993.
- [Yao15] Andrew Chi-Chih Yao, *An n -to-1 bidder reduction for multi-item auctions and its applications*, Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2015, pp. 92–109.
- [Zha08] Yufei Zhao, *Inequalities*, <http://yufeizhao.com/wc08/ineq.pdf>, 2008.

7 Appendix A: Proofs from Section 2

Proof. Proof of Proposition 2.2. By the definition of PBD, the customer can purchase any non-empty subset S of items for the price of $P_0^{\text{PBD}} - \sum_{i \notin S} P_i^{\text{PBD}}$. Of course, the customer can also choose not to make a purchase. Altogether, the class of menus represented by PBD is

$$\left\{ \left(\mathbb{1}_S, P_0^{\text{PBD}} - \sum_{i \notin S} P_i^{\text{PBD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \quad (9)$$

where $\mathbb{1}_S \in \{0, 1\}^n$ is the indicator vector for items belonging to S .

Now, note that (2) defines a valid menu within PCUD since for all i , $P_i^{\text{PCUD}} = P_i^{\text{PBD}} + P_0^{\text{PCUD}} \geq P_0^{\text{PCUD}} = P_0^{\text{PBD}} - \sum_{j=1}^n P_j^{\text{PBD}} \geq 0$. The class of menus represented by (2) is

$$\begin{aligned} & \left\{ \left(\mathbb{1}_S, \sum_{i \in S} P_i^{\text{PCUD}} - (|S| - 1)P_0^{\text{PCUD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \\ &= \left\{ \left(\mathbb{1}_S, \sum_{i \in S} P_i^{\text{PBD}} + (P_0^{\text{PBD}} - \sum_{i=1}^n P_i^{\text{PBD}}) \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \\ &= \left\{ \left(\mathbb{1}_S, P_0^{\text{PBD}} - \sum_{i \notin S} P_i^{\text{PBD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \end{aligned}$$

which is identical to (9). Furthermore, it is easy to see that the relation defined by (2) is a bijection between (9) and

$$\left\{ \left(\mathbb{1}_S, \sum_{i \in S} P_i^{\text{PCUD}} - (|S| - 1)P_0^{\text{PCUD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_i^{\text{PCUD}} \geq P_0^{\text{PCUD}} \geq 0 \ \forall i \in [n] \}.$$

Similarly, note that (3) defines a valid menu within TP since $P_0^{\text{TP}} = P_0^{\text{PBD}} - \sum_{i=1}^n P_i^{\text{PBD}} \geq 0$, and for all i , $P_i^{\text{TP}} = P_i^{\text{PBD}} \geq 0$. The class of menus represented by (3) is

$$\begin{aligned} & \left\{ \left(\mathbb{1}_S, P_0^{\text{TP}} + \sum_{i \in S} P_i^{\text{TP}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \\ &= \left\{ \left(\mathbb{1}_S, (P_0^{\text{PBD}} - \sum_{i=1}^n P_i^{\text{PBD}}) + \sum_{i \in S} P_i^{\text{PBD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \\ &= \left\{ \left(\mathbb{1}_S, P_0^{\text{PBD}} - \sum_{i \notin S} P_i^{\text{PBD}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_1^{\text{PBD}} \geq 0, \dots, P_n^{\text{PBD}} \geq 0, P_0^{\text{PBD}} \geq P_1^{\text{PBD}} + \dots + P_n^{\text{PBD}} \} \end{aligned}$$

which is identical to (9). Furthermore, it is easy to see that the relation defined by (3) is a bijection between (9) and

$$\left\{ \left(\mathbb{1}_S, P_0^{\text{TP}} + \sum_{i \in S} P_i^{\text{TP}} \right) : S \neq \emptyset \right\} \cup \{(0, 0)\} : P_0^{\text{TP}} \geq 0, P_1^{\text{TP}} \geq 0, \dots, P_n^{\text{TP}} \geq 0 \}.$$

This completes the proof of Proposition 2.2. \square

Proof. Proof of Proposition 2.4. Consider any valuation vector $x \in \mathbb{R}^n$. First suppose the customer bought the bundle with all the items for P^{PB} . Under the PBD menu, the customer will still buy the bundle, since it is non-negative utility even if she keeps all the items. However, she will choose to return any items i with $x_i < c_i$. Let S denote the set of such items, which is possibly empty.

- The producer surplus under PB is $P^{\text{PB}} - \sum_{i=1}^n c_i$. The producer surplus under PBD is $(P^{\text{PB}} - \sum_{i \in S} c_i) - \sum_{i \notin S} c_i$, which is identical.
- The consumer surplus under PB is $\sum_{i=1}^n x_i - P^{\text{PB}}$. The consumer surplus under PBD is $\sum_{i \notin S} x_i - (P^{\text{PB}} - \sum_{i \in S} c_i) = \sum_{i=1}^n \max\{x_i, c_i\} - P^{\text{PB}}$ which can only be greater than the consumer surplus under PB.
- The deadweight loss is 0 in both cases: under PB every item is transferred, whereas under PBD every item valued above cost is still transferred.
- The overinclusion loss under PB is $\sum_{i \in S} (c_i - x_i) \geq 0$. The overinclusion loss under PBD is 0, since items in S are not transferred.

On the other hand, suppose the customer did not buy the bundle with all the items for P^{PB} .

- The producer surplus under PB is 0. The producer surplus under PBD is either 0 or $P^{\text{PB}} - \sum_{i=1}^n c_i$ (if the return option allowed the customer to enter the market), which is non-negative.
- The consumer surplus under PB is 0. The consumer surplus under PBD cannot be negative, since the customer is rational and the no-purchase option is always available.

- The deadweight loss under PB is $\sum_{i: x_i > c_i} (x_i - c_i)$, which is the maximum possible. Therefore, the deadweight loss under PBD cannot be greater.
- The overinclusion loss under PB is 0. The overinclusion loss under PBD is always 0 when $P_i^{\text{PBD}} = c_i$ for all i , since items valued below cost are never transferred.

In both cases, we have proven that the statements in Proposition 2.4 hold. \square

Proof. Proof of Proposition 2.5. The firm's problem is to find a menu along with tie-breaking rules which maximize profit. Note that this is equivalent to finding functions q, s defined on \mathcal{X} which are incentive-compatible, individually rational, and profit-maximizing. Formally, the firm's problem is

$$\begin{aligned} \max \quad & \mathbb{E}_{x \sim D}[s(x) - q(x)^T c] \\ \text{s.t.} \quad & q(x)^T x - s(x) \geq q(y)^T x - s(y) \quad \forall x, y \in \mathcal{X} \\ & q(x)^T x - s(x) \geq 0 \quad \forall x \in \mathcal{X} \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \max \quad & \mathbb{E}_{x \sim D}[s(x) - q(x)^T c] \\ \text{s.t.} \quad & q(x)^T (x - c) - (s(x) - q(x)^T c) \geq q(y)^T (x - c) - (s(y) - q(y)^T c) \quad \forall x, y \in \mathcal{X} \\ & q(x)^T (x - c) - (s(x) - q(x)^T c) \geq 0 \quad \forall x \in \mathcal{X} \end{aligned}$$

Now, define $x' := x - c$, $y' := y - c$, $q'(x) := q(x + c)$, and $s'(x) := s(x + c) - q(x + c)^T c$. Let $\mathcal{X}' := \{x - c : x \in \mathcal{X}\}$, and similarly let D' be the distribution D shifted c_i units downward in dimension i for every $i \in [n]$. We can see that the above is equivalent to

$$\begin{aligned} \max \quad & \mathbb{E}_{x' \sim D'}[s'(x')] \\ \text{s.t.} \quad & q'(x')^T x' - s'(x') \geq q'(y')^T x' - s'(y') \quad \forall x', y' \in \mathcal{X}' \\ & q'(x')^T x' - s'(x') \geq 0 \quad \forall x' \in \mathcal{X}' \end{aligned}$$

which is identical to the original problem without costs on this new distribution D' .

Now suppose there was a restriction on the menu $\mathcal{M} = \{(q^{(1)}, s^{(1)}), (q^{(2)}, s^{(2)}), \dots\}$ to belong to some class \mathcal{M} in the original problem. The menu after the transformation, \mathcal{M}' , looks like $\{(q^{(1)}, s^{(1)} - (q^{(1)})^T c), (q^{(2)}, s^{(2)} - (q^{(2)})^T c), \dots\}$. Therefore, \mathcal{M}' is restricted to the class

$$\mathcal{M}' := \{(q^{(1)}, s^{(1)} - (q^{(1)})^T c), (q^{(2)}, s^{(2)} - (q^{(2)})^T c), \dots\} : \{(q^{(1)}, s^{(1)}), (q^{(2)}, s^{(2)}), \dots\} \in \mathcal{M}.$$

By assumption that $s - q^T c \geq 0$ for all menu entries, the payments in \mathcal{M}' are non-negative. \square

Throughout this paper, it will be clear whether we are in the context of the original problem or the transformed problem, and we will omit the superscripts used in the preceding proof.

Remark 7.1. As a concrete example of this transformation, consider the pricing scheme PBDC. \mathcal{M} is restricted to be of the form $\{(\mathbb{1}_S, P_0^{\text{PBD}} - \mathbb{1}_{[n] \setminus S}^T c) : \emptyset \neq S \subseteq [n]\} \cup \{(0, 0)\}$ where $\mathbb{1}_S \in \{0, 1\}^n$ is the indicator vector for items belonging to S . Hence \mathcal{M}' is restricted to be of the form

$$\{(\mathbb{1}_S, P_0^{\text{PBD}} - \mathbb{1}_{[n] \setminus S}^T c - \mathbb{1}_S^T c) : \emptyset \neq S \subseteq [n]\} \cup \{(0, 0)\} = \{(\mathbb{1}_S, P_0^{\text{PBD}} - \mathbb{1}_{[n]}^T c) : \emptyset \neq S \subseteq [n]\} \cup \{(0, 0)\}.$$

Put in words, \mathcal{M}' must belong to the class of menus that offer the same price for any non-empty subset of items. The fact that the customer can choose to take a subset of items instead of taking all the items is important, because valuations x'_i can be negative (x'_i is equal to the original valuation x_i subtract the cost c_i).

8 Appendix B: Proof of Theorem 4.1

We will WOLOG normalize the valuations so that the optimal PC revenue is 1 (we can do this so long as the original optimal revenue was positive; if it was 0 then the statement of the theorem is trivial).

8.1 The Core-Tail Decomposition

We use the core-tail decomposition of [BILW14], with the original idea coming from [LY13]. We will cut up the domain of the joint distribution and consider the conditional distributions on the smaller subdomains. Below, we introduce the notation for working with these distributions on smaller subdomains. One should get comfortable with the idea that some of the distributions defined could be the null distribution, if they were distributions conditioned on a set of measure 0, or a product over an empty set of distributions. The product of a null distribution with any other distribution is still a null distribution.

Definition 8.1. We make the following definitions for this appendix.

- For all $i \in [n]$, let r_i denote the optimal revenue earned by selling item i individually (by our normalization, $\sum_{i=1}^n r_i = 1$).
- Let D_i^C (the “core” of D_i) denote the conditional distribution of D_i when it lies in the range $(-\infty, 1]$.
- Let D_i^T (the “tail” of D_i) denote the conditional distribution of D_i when it lies in the range $(1, \infty)$.
- Let $p_i := \mathbb{P}_{x_i \sim D_i}[x_i > 1]$, the probability item i lies in its tail.
- Let $A \subseteq [n]$ represent a subset of items, usually the items whose valuations lie in their tails.
- Let $D_A^T := \times_{i \in A} D_i^T$, the product distribution of only items in their tails.
- Let $D_A^C := \times_{i \notin A} D_i^C$, the product distribution of only items in their cores.
- Let $D_A := D_A^C \times D_A^T$, the conditional distribution of D when exactly the subset A of items lie in their tails. Let p_A be the probability this occurs, which is equal to $(\prod_{i \notin A} (1 - p_i))(\prod_{i \in A} p_i)$, by independence.
- Let $x_i^+ := \max\{x_i, 0\}$.

- For any valuation distribution S , let $\text{VAL}^+(S) := \sum_i \mathbb{E}_{x \sim S}[x_i^+]$, which is the expected welfare after the transformation from costs to negative valuations. Note that the sum is only over the admissible i if S is a distribution on a smaller subdomain.
- Let $\text{REV}(S)$ denote the optimal revenue obtainable from valuation distribution S via any Incentive Compatible and Individually Rational mechanism, which could include lotteries.
- Let $\text{SREV}(S)$ denote the optimal revenue of any pricing scheme falling under the class of separate sales (Pure Components).
- Let $\text{BDCREV}(S)$ denote the optimal revenue of any pricing scheme falling under the class of PBDC.

(It is understood that $\text{VAL}^+, \text{REV}, \text{SREV}, \text{BDCREV}$ are 0 when evaluated on the null distribution.)

8.2 Lemmas for Negative Valuations

We need to modify the statements of lemmas from [HN12], [LY13], and [BILW14] to handle negative valuations. While their proofs can be extended to negative valuations in a straight-forward manner, we provide full self-contained proofs here for ease of exposition.

Lemma 8.2. (*Marginal Mechanism*) *Let S, S' be (potentially negative) valuation distributions over disjoint sets of items. Then*

$$\text{REV}(S \times S') \leq \text{VAL}^+(S) + \text{REV}(S')$$

The Marginal Mechanism tells us that when selling a group of independent items, we cannot do better than breaking off some items individually, extracting the entire welfare from those items, and selling the remaining items as a group.

Proof. Proof of Lemma 8.2. Consider the following mechanism for selling to a buyer with valuations drawn from S' . First, sample a value $v \sim S$, and reveal to the buyer these make-believe valuations for the items in S . Then run a mechanism obtaining $\text{REV}(S \times S')$ on this buyer, with the modification that whenever the buyer would have received an item i from the support of S , instead she will receive (or pay) money equal to v_i . By independence, this modified mechanism on the buyer with valuations drawn from S' is IC and IR (a buyer with valuations S' will choose the same menu entry under the modified mechanism as a buyer with valuations $S \times S'$ would have chosen under the original mechanism) and we will obtain $\text{REV}(S \times S')$, but then have to settle for the items in S . The most we stand to lose in the settlement is $\sum_i v_i^+$ (each item i in S is transferred in full whenever $v_i \geq 0$, and not transferred when $v_i < 0$), so this amount is upper bounded in expectation by $\text{VAL}^+(S)$. Therefore, the optimal revenue from S' is at least $\text{REV}(S \times S') - \text{VAL}^+(S)$, completing the proof of the lemma. \square

Lemma 8.3. (*Subdomain Stitching*) Let S be a product distribution over valuations, with support $\mathcal{X} \subseteq \mathbb{R}^m$ for some $m \in \mathbb{N}$. Let $\mathcal{X}_1, \dots, \mathcal{X}_k$ form a partition of \mathcal{X} inducing conditional distributions $S^{(1)}, \dots, S^{(k)}$, respectively, and let $s_j = \mathbb{P}_{x \sim S}[x \in \mathcal{X}_j]$. Then

$$\text{REV}(S) \leq \sum_{j=1}^k s_j \text{REV}(S^{(j)})$$

Intuitively, Subdomain Stitching says that revenue can only increase if we sell to each subdomain separately, since we can use a different mechanism for each subdomain that specializes in extracting the welfare from that customer segment.

Proof. Proof of Lemma 8.3. Let M be an optimal mechanism obtaining $\text{REV}(S)$, and for any valuation distribution S' , let $\text{REV}_M(S')$ denote the expected revenue obtained from mechanism M when the buyer's valuation is drawn from S' . Clearly $\text{REV}(S) = \sum_{j=1}^k s_j \text{REV}_M(S^{(j)})$, and furthermore for all $j \in [k]$, $\text{REV}_M(S^{(j)}) \leq \text{REV}(S^{(j)})$ since M is an IC-IR mechanism for selling to $S^{(j)}$, completing the proof of the lemma. \square

Lemma 8.4. Let S be a product distribution over valuations, with support $\mathcal{X} \subseteq \mathbb{R}^m$ for some $m \in \mathbb{N}$. Let \mathcal{X}' be a subset of \mathcal{X} inducing conditional distribution S' , and let $s' = \mathbb{P}_{x \sim S}[x \in \mathcal{X}']$. Then

$$\text{REV}(S) \geq s' \text{REV}(S')$$

While Subdomain Stitching places an upper bound on $\text{REV}(S)$, Lemma 8.4 places a lower bound on $\text{REV}(S)$ based on the optimal revenue of any single subdomain.

Proof. Proof of Lemma 8.4. Consider an optimal mechanism for S' , and extend this to an IC-IR mechanism on S by allowing the buyer to report a value in \mathcal{X}' maximizing her utility. With probability s' , the buyer's valuation will actually be drawn from S' and we will obtain revenue $\text{REV}(S')$; otherwise, we still earn a non-negative revenue, since the mechanism never admits a negative payment. Therefore, the optimal revenue for S is at least $s' \text{REV}(S')$, completing the proof of the lemma. \square

Lemma 8.5. Let S be a product distribution over m independent (potentially negative) valuations, for some $m \in \mathbb{N}$. Then

$$\text{REV}(S) \leq m \cdot \text{SREV}(S)$$

While selling m items together can definitely be better than selling them separately, this lemma tells us it can be no more than m times better.

Proof. Proof of Lemma 8.5. We proceed by induction. The statement is trivial when $m = 1$. Now, suppose we have proven the statement for m valuations, and we will prove it for $m + 1$ valuations.

Partition the support $\mathcal{X} \subseteq \mathbb{R}^{m+1}$ of S into \mathcal{X}_1 and \mathcal{X}_2 , where $\mathcal{X}_1 := \{x \in \mathcal{X} : x_1 \geq \max\{x_j, 0\} \forall j = 2, \dots, m+1\}$ and $\mathcal{X}_2 := \mathcal{X} \setminus \mathcal{X}_1$. Let s_1 denote the probability a value sampled from S lies in \mathcal{X}_1 , and let S_1 be its distribution conditioned on this event. Define s_2, S_2 respectively. Subdomain

stitching tells us $\text{REV}(S) \leq s_1 \text{REV}(S^{(1)}) + s_2 \text{REV}(S^{(2)})$. Our goal is to separately show that $s_1 \text{REV}(S^{(1)}) \leq (m+1) \text{SREV}(S_1)$ and $s_2 \text{REV}(S^{(2)}) \leq (m+1) \text{SREV}(S_{-1})$.

Now, applying Marginal Mechanism on $S^{(1)}$ and multiplying both sides of the inequality by s_1 , we get $s_1 \text{REV}(S^{(1)}) \leq s_1 \text{VAL}^+(S_{-1}^{(1)}) + s_1 \text{REV}(S_1^{(1)})$. By considering a distribution that samples $v \sim S$ but only outputs v_1 , we can use Lemma 8.4 to show that $s_1 \text{REV}(S_1^{(1)}) \leq \text{REV}(S_1)$. To bound $\text{VAL}^+(S_{-1}^{(1)})$, consider the following mechanism for selling just item 1: sample $v_{-1} \sim S_{-1}$, and set the price to be $\max_{i=2}^{m+1} \{\max\{v_i, 0\}\}$. Since the buyer's valuation is drawn from S_1 , by independence, we get a sale with probability exactly s_1 . Furthermore, $\max_{i=2}^{m+1} \{\max\{v_i, 0\}\} \geq \frac{1}{m} \sum_{i=2}^{m+1} \max\{v_i, 0\}$, so conditioned on us getting a sale, the expected payment is at least $\frac{1}{m} \text{VAL}^+(S_{-1}^{(1)})$. We have proven $\text{REV}(S_1) \geq \frac{s_1}{m} \text{VAL}^+(S_{-1}^{(1)})$, hence $s_1 \text{REV}(S^{(1)}) \leq (m+1) \text{REV}(S_1) = (m+1) \text{SREV}(S_1)$, as required.

It remains to bound $s_2 \text{REV}(S^{(2)})$, and using Marginal Mechanism and Lemma 8.4 in the same way as before, we obtain that it is no more than $s_2 \text{VAL}^+(S_1^{(2)}) + \text{REV}(S_{-1})$. Consider the following mechanism for selling items $2, \dots, m+1$: sample $v_1 \sim S_1$, and set the individual price for each item $2, \dots, m+1$ to be $\max\{v_1, 0\}$. Note that the probability of getting at least one sale is less than s_2 , since even when there is some $j = 2, \dots, m+1$ such that $v_1 < \max\{x_j, 0\}$, it is possible for both v_1, x_j to be negative. However, in this case $\max\{v_1, 0\} = 0$, so not getting a sale is still equivalent to getting at least one sale for $\max\{v_1, 0\}$. Therefore, we can think of it as we get at least one sale with probability s_2 , in which case we earn in expectation at least $\text{VAL}^+(S_1^{(2)})$. We have proven that $s_2 \text{VAL}^+(S_1^{(2)}) \leq \text{SREV}(S_{-1})$, and by the induction hypothesis $\text{REV}(S_{-1}) \leq m \cdot \text{SREV}(S_{-1})$, so $s_2 \text{REV}(S^{(2)}) \leq (m+1) \text{SREV}(S_{-1})$.

Putting everything together, we have $\text{REV}(S) \leq (m+1)(\text{SREV}(S_1) + \text{SREV}(S_{-1})) = (m+1) \text{SREV}(S)$, completing the induction and the proof of the lemma. \square

Using these lemmas, we decompose the revenue of the initial distribution D in the same way as [BILW14]:

$$\begin{aligned}
\text{REV}(D) &\leq \sum_{A \subseteq [n]} p_A \text{REV}(D_A) \\
&\leq \sum_{A \subseteq [n]} p_A (\text{VAL}^+(D_A^C) + \text{REV}(D_A^T)) \\
&\leq \sum_{A \subseteq [n]} p_A \text{VAL}^+(D_\emptyset^C) + \sum_{A \subseteq [n]} p_A \text{REV}(D_A^T) \\
&= \text{VAL}^+(D_\emptyset^C) + \sum_{A \subseteq [n]} p_A \text{REV}(D_A^T)
\end{aligned}$$

where the first inequality is Subdomain Stitching, the second inequality is Marginal Mechanism, the third inequality is immediate from the definition of D_A^C , and the equality is a consequence of $\sum_{A \subseteq [n]} p_A = 1$.

Now, for all $A \subseteq [n]$ such that $p_A > 0$, Lemma 8.5 tells us that $\text{REV}(D_A^T) \leq |A| \text{SREV}(D_A^T) =$

$|A| \sum_{i \in A} \text{SREV}(D_i^T)$. Lemma 8.4 tells us that $\text{SREV}(D_i^T) \leq \frac{r_i}{p_i}$, where $p_i \neq 0$ since $p_A > 0$, so

$$\begin{aligned} \sum_{A \subseteq [n]} p_A \text{REV}(D_A^T) &\leq \sum_{A \subseteq [n]} p_A |A| \sum_{i \in A} \frac{r_i}{p_i} \\ &= \sum_{i=1}^n r_i \sum_{A \ni i} |A| \frac{p_A}{p_i} \end{aligned}$$

$\sum_{A \ni i} |A| \frac{p_A}{p_i}$ is the expected number of items in their tails conditioned on item i being in its tail, so it is equal to $1 + \sum_{j \neq i} p_j$. Thus

$$\begin{aligned} \sum_{A \subseteq [n]} p_A \text{REV}(D_A^T) &\leq \sum_{i=1}^n r_i \left(1 + \sum_{j \neq i} p_j\right) \\ &= 1 + \sum_{j=1}^n p_j \sum_{i \neq j} r_i \\ &= 1 + \sum_{j=1}^n p_j (1 - r_j) \end{aligned}$$

We will use τ to denote the quantity $\sum_{i=1}^n p_i (1 - r_i)$. It is immediate that $\tau \leq \sum_{i=1}^n p_i \leq 1$, but we can get a stronger bound for the welfare of the core if we don't immediately apply the inequality $\tau \leq 1$. We have

$$\text{REV}(D) \leq \text{VAL}^+(D_\emptyset^C) + 1 + \tau \quad (10)$$

Before we proceed, one final lemma we will need later is:

Lemma 8.6. *Let Y be a random variable distributed over $[0, 1]$ and suppose $y(1 - F(y))$ is upper bounded by some value $v \in [0, 1]$. Then $\text{Var}(Y) \leq 2v$.*

Proof. Proof of Lemma 8.6.

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &\leq \mathbb{E}[Y^2] \\ &= \int_0^1 \mathbb{P}[Y^2 \geq y] dy \\ &= \int_0^1 \mathbb{P}[Y \geq \sqrt{y}] dy \\ &\leq \int_0^1 \frac{v}{\sqrt{y}} dy \\ &= 2v \end{aligned}$$

where the second inequality uses the fact that the Myerson revenue for Y is upper bounded by v . \square

8.3 A Tighter Bound for the Welfare of the Core

The main observation behind our improvement is that for τ to be large (and the above bound to be weak), the tail probabilities must be large. However, we will choose the price of the grand bundle, P_t , to be at most 2, so that whenever 2 or more valuations lie in their tails, the customer is guaranteed to want to buy the bundle (and dispose of items for which her valuation is negative). Thus

$$\begin{aligned}
\mathbb{P}[\sum x_i^+ < P_t] &= p_\emptyset \cdot \mathbb{P}_{x \sim D_\emptyset}[\sum x_i^+ < P_t] + \sum_{|A|=1} p_A \cdot \mathbb{P}_{x \sim D_A}[\sum x_i^+ < P_t] + \sum_{|A| \geq 2} p_A \cdot (0) \\
&\leq \left(p_\emptyset + \sum_{|A|=1} p_A \right) \cdot \mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t] \\
&= \left(\prod_{i=1}^n (1 - p_i) + \sum_{i=1}^n p_i \prod_{j \neq i} (1 - p_j) \right) \cdot \mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t]
\end{aligned} \tag{11}$$

where the inequality comes from the fact that the probability of $\sum x_i^+$ being less than the bundle price is greater conditioned on no items being in the tail, than conditioned on some item being in the tail. We used independence to compute the probabilities in the final expression, which we will bound in the following way:

Lemma 8.7. *Let $p_1, \dots, p_n, r_1, \dots, r_n$ be real numbers satisfying $0 \leq p_i \leq r_i$ and $\sum_{i=1}^n r_i = 1$. Let $\tau = \sum_{i=1}^n p_i(1 - r_i)$. Then*

$$\prod_{i=1}^n (1 - p_i) + \sum_{i=1}^n p_i \prod_{j \neq i} (1 - p_j) \leq \frac{\frac{5}{4} + \tau}{e^\tau}$$

This is the key inequality that enables our improved ratio and its proof requires new analysis. Note that we do indeed have the condition $p_i \leq r_i$ in our case, since by Lemma 8.4 $r_i \geq p_i \text{REV}(D_i^T)$, and $\text{REV}(D_i^T)$ must be at least 1 when D_i^T is distributed over $(1, \infty)$.

Proof. Proof of Lemma 8.7. We will first prove

$$\frac{3}{4} \cdot \prod_{i=1}^n (1 - p_i) + \sum_{i=1}^n p_i \prod_{j \neq i} (1 - p_j) \leq \frac{1 + \tau}{e^\tau} \tag{12}$$

Assume that $p_i < 1$ for all $i \in [n]$; the lemma is trivially true otherwise because we would have LHS = 1 and $\tau = 0$. Since $\tau = \sum_{i=1}^n p_i(1 - r_i)$ and $1 - x \leq e^{-x}$, it suffices to prove

$$\frac{3}{4} \cdot \prod_{i=1}^n (1 - p_i) + \sum_{i=1}^n p_i \prod_{j \neq i} (1 - p_j) \leq \left(1 + \sum_{i=1}^n p_i(1 - r_i) \right) \prod_{i=1}^n (1 - p_i(1 - r_i))$$

which is equivalent to

$$\frac{3}{4} + \sum_{i=1}^n \frac{p_i}{1 - p_i} \leq \left(1 + \sum_{i=1}^n (p_i - p_i r_i) \right) \prod_{i=1}^n \left(1 + \frac{p_i r_i}{1 - p_i} \right)$$

Observe that the RHS is at least

$$\begin{aligned}
& \left(1 + \sum_{i=1}^n (p_i - p_i r_i)\right) \left(1 + \sum_{i=1}^n \frac{p_i r_i}{1 - p_i}\right) \\
&= 1 + \sum_{i=1}^n \frac{(p_i - p_i r_i)(1 - p_i) + p_i r_i}{1 - p_i} + \left(\sum_{i=1}^n p_i(1 - r_i)\right) \left(\sum_{i=1}^n \frac{p_i r_i}{1 - p_i}\right) \\
&= 1 + \sum_{i=1}^n \frac{p_i}{1 - p_i} - \sum_{i=1}^n \frac{p_i^2(1 - r_i)}{1 - p_i} + \left(\sum_{i=1}^n p_i(1 - r_i)\right) \left(\sum_{i=1}^n \frac{p_i r_i}{1 - p_i}\right) \\
&= 1 + \sum_{i=1}^n \frac{p_i}{1 - p_i} - \sum_{i=1}^n \frac{p_i^2(1 - r_i)^2}{1 - p_i} + \sum_{i \neq j} p_i(1 - r_i) \cdot \frac{p_j r_j}{1 - p_j}
\end{aligned}$$

so it remains to prove

$$\sum_{i=1}^n \frac{p_i^2(1 - r_i)^2}{1 - p_i} - \sum_{i \neq j} p_i(1 - r_i) \cdot \frac{p_j r_j}{1 - p_j} \leq \frac{1}{4}$$

But $p_i \leq r_i$ for all $i \in [n]$, so the LHS is at most $\sum_{i=1}^n p_i^2(1 - p_i)$, which can be seen to be at most $\frac{1}{4}$, since $p_i(1 - p_i)$ is always at most $\frac{1}{4}$ and $\sum_{i=1}^n p_i \leq 1$.

Also, since $\tau \leq \sum_{i=1}^n p_i$, $e^{-\tau} \geq \exp(-\sum_{i=1}^n p_i) \geq \prod_{i=1}^n (1 - p_i)$. Multiplying by $\frac{1}{4}$ and adding to (12), we complete the proof of the lemma. \square

8.4 Applying Cantelli's Inequality

To bound $\mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t]$, we want to show that $\sum x_i^+$ concentrates around its mean, where valuation x_i is drawn from its conditional core distribution D_i^C for all $i \in [n]$. Note that $y(1 - F_{x_i}(y))$ is bounded above by r_i for all $y \in [0, 1]$; otherwise $\text{SREV}(D_i^C) > r_i \implies \text{SREV}(D_i) > r_i$ which is a contradiction. Hence $y(1 - F_{x_i^+}(y))$ is also bounded above by r_i and we can invoke Lemma 8.6 to get $\text{Var}_{x_i \sim D_i^C}(x_i^+) \leq 2r_i$ for all $i \in [n]$. By independence, $\text{Var}_{x \sim D_\emptyset^C}(\sum x_i^+) = \sum_{i=1}^n \text{Var}_{x \sim D_\emptyset^C}(x_i^+) \leq \sum_{i=1}^n 2r_i = 2$ and we have successfully bounded the variance of the quantity we are interested in.

At this point, it is common in the literature to see an application of Chebyshev's inequality (e.g. [BB99, FN06, HN12, BILW14]). However, since we are only interested in the lower tail, we can actually use Cantelli's one-sided inequality (Lemma 3.1), which optimizes a shift parameter to obtain an improved bound for a single tail.

Now, note that $\mathbb{E}_{x \sim D_\emptyset^C}[\sum_{i=1}^n x_i^+] = \text{VAL}^+(D_\emptyset^C)$ by definition. Also, it will be convenient to write the bundle price as $P_t = \alpha \cdot \text{VAL}^+(D_\emptyset^C)$, for some $\alpha \in [0, 1]$ (we would never want $\alpha > 1$ since then

the price would be greater than the mean and it would be impossible to use Cantelli). Then

$$\begin{aligned}
\mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t] &= \mathbb{P}_{x \sim D_\emptyset^C} \left[\sum_{i=1}^n x_i^+ - \text{VAL}^+(D_\emptyset^C) < -(1-\alpha)\text{VAL}^+(D_\emptyset^C) \right] \\
&\leq \frac{\text{Var}_{x \sim D_\emptyset^C}(\sum x_i^+)}{\text{Var}_{x \sim D_\emptyset^C}(\sum x_i^+) + (1-\alpha)^2 \text{VAL}^+(D_\emptyset^C)^2} \\
&\leq \frac{2}{2 + (1-\alpha)^2 \text{VAL}^+(D_\emptyset^C)^2}
\end{aligned}$$

where the first inequality is Cantelli's inequality, and the second inequality comes from our variance bound above. So long as we choose $P_t \leq 2$, we can use (11), and combined with Lemma 8.7 we get

$$\mathbb{P}[\sum x_i^+ < P_t] \leq \min \left\{ \frac{1.25 + \tau}{e^\tau}, 1 \right\} \cdot \frac{2}{2 + (1-\alpha)^2 \text{VAL}^+(D_\emptyset^C)^2}$$

and hence the expected revenue from selling the grand bundle at price $\alpha \cdot \text{VAL}^+(D_\emptyset^C)$ is at least

$$\alpha \cdot \text{VAL}^+(D_\emptyset^C) \cdot \left(1 - \min \left\{ \frac{1.25 + \tau}{e^\tau}, 1 \right\} \cdot \frac{2}{2 + (1-\alpha)^2 \text{VAL}^+(D_\emptyset^C)^2} \right)$$

Recall from (10) that $\text{REV}(D) \leq \text{VAL}^+(D_\emptyset^C) + 1 + \tau$. While τ could take on any value in $[0, 1]$, we can choose the price of the bundle based on τ and $\text{VAL}^+(D_\emptyset^C)$ by adjusting $\alpha \in [0, 1]$.

Case 1. If $\text{VAL}^+(D_\emptyset^C) \leq 3.2$, then $\text{REV}(D) \leq 3.2 + 1 + 1 = 5.2 \cdot \text{SREV}(D)$ is immediate and we can just sell the items individually.

Case 2. If $3.2 < \text{VAL}^+(D_\emptyset^C) \leq 4$, then we will choose $\alpha = \frac{1}{2}$ which guarantees $P_t \leq 2$. Thus

$$\text{BDCREV}(D) \geq \text{VAL}^+(D_\emptyset^C) \cdot \frac{1}{2} \left(1 - \min \left\{ \frac{1.25 + \tau}{e^\tau}, 1 \right\} \cdot \frac{2}{2 + (1 - \frac{1}{2})^2 (3.2)^2} \right)$$

It can be shown with calculus (or numerically) that:

Proposition 8.8. For all $\tau \in [0, 1]$, $2 \left(1 - \min \left\{ \frac{1.25 + \tau}{e^\tau}, 1 \right\} \cdot \frac{2}{2 + (1 - \frac{1}{2})^2 (3.2)^2} \right)^{-1} + (1 + \tau) < 5.2$, with the maximum of ≈ 5.1952 occurring at the unique positive τ satisfying $\frac{1.25 + \tau}{e^\tau} = 1$.

Hence $\text{VAL}^+(D_\emptyset^C) \leq (4.2 - \tau) \text{BDCREV}(D)$. Substituting into (10), we get

$$\begin{aligned}
\text{REV}(D) &\leq (4.2 - \tau) \text{BDCREV}(D) + (1 + \tau) \text{SREV}(D) \\
&\leq 5.2 \cdot \max\{\text{SREV}(D), \text{BDCREV}(D)\}
\end{aligned}$$

as desired.

Case 3. If $4 < \text{VAL}^+(D_\emptyset^C)$, then we will still choose $\alpha = \frac{1}{2}$. We no longer have $P_t \leq 2$, so we have to use the weaker bound $\mathbb{P}_{x \sim D}[\sum x_i^+ < P_t] \leq \mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t]$. However, applying Cantelli yields

$$\mathbb{P}_{x \sim D_\emptyset^C}[\sum x_i^+ < P_t] \leq \frac{2}{2 + (1 - \frac{1}{2})^2 (4)^2} = \frac{1}{3}$$

so $\text{BDCREV}(D) \geq \text{VAL}^+(D_\emptyset^C) \cdot \frac{1}{2} (1 - \frac{1}{3})$. We get $\text{REV}(D) \leq 3 \cdot \text{BDCREV}(D) + (1 + \tau) \text{SREV}(D) < 5.2 \cdot \max\{\text{SREV}(D), \text{BDCREV}(D)\}$, completing the proof of Theorem 4.1.

9 Appendix C: Proof of Theorem 4.3

It is immediate that the optimal revenue from PC is 2ρ , attained by selling individual items at any price in $[1, 2]$. Next, we would like to argue that the optimal revenue from PB is also 2ρ . If we offer the bundle at 2, it is guaranteed to get bought if either valuation realizes to 2 or both valuations realize to a positive number, and won't get bought otherwise. Therefore the revenue is $2(\rho^2 + 2(1 - \rho)\frac{\rho}{2}) = 2\rho$.

We can do equally well by offering the bundle at 3, and any other price is inferior.

Lemma 9.1. *The optimal revenue from PB is 2ρ , attained by setting a bundle price of 2 or 3.*

Proof. Proof of Lemma 9.1. Let z denote the price of the bundle. We will systematically analyze all the cases over $1 \leq z \leq 4$ and show that the maximum revenue of 2ρ is attained at $z = 2$ and $z = 3$.

Case 1. Suppose $1 \leq z \leq 2$. Let us condition on the realization y of the first valuation. If $y = 0$, then we get a sale with probability $\frac{\rho}{z}$. If $y \in [1, z)$, then we get a sale so long as the second valuation realizes to a positive number, which occurs with probability $1 - \rho$. If $y \geq z$, then the first valuation alone is enough to guarantee a bundle sale. The expected revenue is

$$z \left((1 - \rho) \frac{\rho}{z} + (\rho - \frac{\rho}{z}) \rho + \frac{\rho}{z} \right) = 2\rho + (z - 2)\rho^2$$

which is clearly maximized at $z = 2$, in which case the revenue is 2ρ .

Case 2. Suppose $2 < z \leq 3$. Let us condition on the realization y of the first valuation. If $y = 0$, then we have no chance of selling the bundle. If $y \in [1, z - 1]$, then we get a sale when the other valuation is at least $z - y$. Since $z - y \in [1, 2]$, the probability of this occurring is $\frac{\rho}{z - y}$. If $y \geq z - 1$, then we get a sale so long as the other valuation realizes to a positive number, which occurs with probability ρ . The total probability of getting a sale is

$$\int_1^{z-1} \frac{\rho}{y^2} \frac{\rho}{z - y} dy + \frac{\rho}{z - 1} \rho$$

where the PDF of Y satisfies $f(y) = \frac{\rho}{y^2}$ over $[1, 2]$. Using partial fractions, the antiderivative of $\frac{1}{y^2(z - y)}$ can be computed to be

$$\frac{1}{z} \left(\frac{\ln y - \ln(z - y)}{z} - \frac{1}{y} \right)$$

as demonstrated in the proof of [HN12, lem. 6]. Therefore, the definite integral evaluates to

$$\rho^2 \left(\frac{2 \ln(z - 1)}{z^2} + \frac{2}{z} - \frac{1}{z - 1} \right)$$

and the expected revenue is

$$z\rho^2 \left(\frac{2 \ln(z - 1)}{z^2} + \frac{2}{z} - \frac{1}{z - 1} + \frac{1}{z - 1} \right) = 2\rho^2 \left(\frac{\ln(z - 1)}{z} + 1 \right)$$

However, $\frac{\ln(z - 1)}{z}$ is a strictly increasing function on $(2, 3]$, so this expression is uniquely maximized at $z = 3$ where it equals $2\rho^2(\frac{\ln 2}{3} + 1) = 2\rho$.

Case 3. Suppose $3 \leq z \leq 4$. Let us condition on the realization y of the first valuation. If $y < z - 2$, then we have no chance of selling the bundle. Otherwise, the probability of getting a sale is $\frac{\rho}{z-y}$, since $z - y \in [1, 2]$. The total probability of getting a sale is

$$\int_{z-2}^2 \frac{\rho}{y^2} \frac{\rho}{z-y} + \frac{\rho}{2} \frac{\rho}{z-2}$$

and the integral evaluates to

$$\rho^2 \left(\frac{2 \ln 2 - 2 \ln(z-2)}{z^2} + \frac{1}{z(z-2)} - \frac{1}{2z} \right)$$

Therefore, the expected revenue is

$$z\rho^2 \left(\frac{2 \ln 2 - 2 \ln(z-2)}{z^2} + \frac{1}{z(z-2)} - \frac{1}{2z} + \frac{1}{2(z-2)} \right) = 2\rho^2 \left(\frac{\ln 2 - \ln(z-2)}{z} + \frac{1}{z-2} \right)$$

$\frac{\ln 2 - \ln(z-2)}{z} + \frac{1}{z-2}$ is a strictly decreasing function on $[3, 4]$, so this expression is uniquely maximized at $z = 3$. \square

Now, consider the strategy of offering either item for 2 or the bundle for the discounted price of 3. Note that if buying the bundle is non-negative utility for the customer, then buying either individual item cannot be higher utility, since the price savings is one and the value of the item lost is at least one (recall that the firm gets to break ties in a way that favors itself). Hence there is no cannibalization of bundle sales from individual sales and we earn revenue at least 2ρ . However, when exactly one valuation realizes to a positive number (in which case we have no chance of selling the bundle), we still have a $\frac{1}{2}$ conditional probability of selling that individual item. Hence the revenue from Mixed Bundling is $2\rho + 2(2(1-\rho)\frac{\rho}{2}) = 2\rho(2-\rho)$.

The relative gain over both the PC revenue and the PB revenue is $2-\rho = \frac{3+2\ln 2}{3+\ln 2}$, completing the proof of Theorem 4.3.

Remark 9.2. A motivating example for our construction is a small modification of the earlier best-known example from [HN12]: consider a distribution that takes on values 0, 1, 2 with probabilities $\frac{1}{9}, \frac{4}{9}, \frac{4}{9}$, respectively. Let D be the instance consisting of two independent copies of this distribution. Then it can be shown that the optimal PC revenue is $\frac{16}{9}$ (attained at individual prices 1 or 2), the optimal PB revenue is $\frac{16}{9}$ (attained at bundle price 2 or 3), and the optimal revenue is at least $\frac{160}{81}$ (attained at individual prices 2 and bundle price 3), achieving a ratio of $\frac{9}{10}$. [HN12] had the probabilities be $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ instead, achieving a ratio of $\frac{12}{13}$.

10 Appendix D: Example where BSP Performs Poorly

Consider a firm that is bundling a higher-profit-margin, lower-valuation good with a low-profit-margin, high-valuation good. This is a common occurrence, for example when video games are bundled with a console, which we will hereinafter refer to as item 1 and item 2, respectively. Item 1 costs zero to produce and has a valuation uniform on $[0,1]$; item 2 costs 4.5 to produce and has

a valuation uniform on $[0,5]$ and independent from item 1. Most of the welfare comes from the lower-valuation item: the expected welfare for item 1 and item 2 are 0.5 and 0.025, respectively.

The optimal deterministic profit is ≈ 0.265 , attained by offering item 1 at 0.51, item 2 at 4.83, and the bundle at the discounted price of 5.13.

The optimal BSP pricing charges 4.83 for a single item and 5.03 for both items, earning only 19% of the deterministic optimum. This example highlights the issue with BSP: it cannot afford to charge a low price for a single item if any item has a high production cost. However, most of the potential profit could be coming from offering lower-valuation items at low prices! [CLS08] bypass such examples in their numerical experiments, assuming that all items have a low cost compared to its mean valuation.

PBDC offers item 1 at 0.51, item 2 at 5.01, and the bundle at 5.01—which is the right idea and earns 99.1% of the deterministic optimum. Interestingly, even the analytical solution provided by [Bha13], which computes the optimal deterministic pricing when there are two independent uniform distributions and costs, is less effective than PBDC on this example. The solution from [Bha13] only attains 97.5% of the deterministic optimum for this example, because it requires a bit of linear approximation.

Optimal bundling is an intricate problem even in the case of two independent uniform distributions, so a simple pricing heuristic as robust as PBDC is invaluable. In fact, for this example PBDC recommends *Partial Mixed Bundling*, which is a Mixed Bundling scheme where one of the items, in this case item 2 (the high-cost low-welfare item), is never sold individually. This matches the intuition that the seller should add item 1 (the low-cost high-welfare item) to item 2 in order to increase the total amount customer is willing to pay (see Proposition 1 in [Bha13]). BSP, on the other hand, does not perform well: it recommends a Partial Mixed Bundling scheme where item 1 is never sold individually.